Role of the C-Terminal Domain of the Aminoacyl-tRNA *Trans*-editing Enzyme ProXp-ala in *Arabidopsis thaliana*

Undergraduate Research Thesis

Presented in partial fulfillment of the requirements for graduation "with Honors Research Distinction in Biochemistry" in the undergraduate colleges of The Ohio State University

by
John Alan Vu

The Ohio State University
April 14, 2021

Project Advisor: Dr. Karin Musier-Forsyth, Department of Chemistry & Biochemistry

Table of Contents

Acknowledgements

There are so many people I have to thank for this work…

First, thank you, Karin, for responding to my very first email inquiry about undergraduate research positions and your decision to invest in me as a student, a scientist, and an individual. Over the past few years, I have grown significantly in all of these aspects, and my time in your lab has contributed so much to my college experience, professional future, and personal development. The dedication that you have showed me in just the past few days polishing this thesis are a tribute to the scientist you are and that I strive to be. I look forward to continuing to work with you and look forward to keeping in touch after I move on to the next chapter of my life. I will never forget my scientific roots, and I am so grateful to have had you for a mentor.

Next, thank you to Jun-Kyu Byun and Lexie Kuzmishin Nagy for their mentorship and comradery in lab. Jun-Kyu, it has been such a pleasure being able to learn about and study this protein with you. Each finding brings us closer and closer to a publication, and I am looking forward to being able to share our hard work with the scientific community and the world. Lexie, you were my scientific and technical foundation for so long, and I would not be the scientist I am today without your training. Thank you for sacrificing so much of your own time to teach me, mold me, and answer all of my questions, no matter how many times I needed them to be repeated. I hope to be able to mentor students of my own one day with the same patience and passion that you showed me.

To the other members of the KMF lab, namely Marina, Danni, Will, Irina, Shuohui, Yu-Ci, and Jawad, thank you all so much for your friendship and your willingness to mentor at help me with the simplest request. You all made coming to lab something I always looked forward to, and I am so thankful to have such a great community of scientists to call my lab family.

To my colleagues in the Simonović lab at UIC, Miljan, Kaitlyn, and Anu, thank you for welcoming me into your lab and making it feel like a lab home away from lab home. I loved being able to exchange tips and tricks and experiences, and I hope to be able to continue to keep in touch with you all as I begin to make a career of my own.

Thank you to Drs. Gopalan and Jang for accepting roles on my thesis committee. I could not ask for more qualified individuals to judge and review my work. I am so grateful to hear your input and look forward to becoming a better scientist as a result of it.

Thank you to all of the institutions that have supported my research career, namely the Department of Chemistry and Biochemistry, the Department of Molecular Genetics, the College of Arts and Sciences, and the Office of Undergraduate Research and Creative Inquiry at Ohio State.

To all of my professors here at OSU especially in the Department of Chemistry and Biochemistry, it has been an honor and privilege to study under such incredible minds. I only hope to use what you have taught me to help chew away at the edge of the unknown, inform the next generations of where we are in our journey for greater understanding, and inspire those future scientists to ask questions about this incredible and mysterious universe, as you have.

Thank you to all of my wonderful friends who have supported and stood by me throughout my college career. You all push me to keep asking questions, be curious, and find joy in both the known and unknown. You have put up with my demanding life as a researcher and still always asked me to hangout no matter how many times I told you "Sorry, I can't hang, I am going to lab…" College is not easy, and you all kept me sane and became my family during this long, challenging, but fun four-year journey. I could not imagine being the person I am today without having you in my life. You know who you are.

Lastly, and most importantly, thank so much to my amazing, loving, and supportive family. Mom, Dad, James, and Jacob, you guys are everything to me, and I want you to know that everything I do comes from you. My time in college has allowed me to grow, explore, and develop my own identity, but I know that it is you guys that keep me grounded no matter how many states lie between us.

James, Jacob, as your older brother, I can't help but want to watch over and protect you like I did when you were so young. Now that you are older, I want you to know that I am so grateful of the men that you have become, and I thank you for always reminding me of my imperfections and for pushing me to constantly grow and be a better version of myself. You are my heroes, and I wish nothing but the world for you both.

Mom, Dad, thank you for teaching me what it means to be strong in mind, body, and spirit. I only hope to one day be as kind, loving, and supportive to my children as you are to me, James, and Jacob. I look to you as examples of how to be a functioning adult (which I am finding is not everything that it is cracked up to be!), and I know that even as I grow to be more independent, you will still love me and welcome me home unconditionally, and I will always be grateful for that. As I begin to make my own way through life as a scientist, I know that I will not always be able to count on you for all the answers, but I know that I will always be able to turn to you for wisdom and love. Thank you for everything.

I probably missed someone, and to them, I am sorry! I just want to stress that though my name is on the title slide of this work, I am a terrible teammate because this work was only possible because of the many wonderful people in my life that pushed me to continue my work for the past four years. You truly do not know your impact, and for that, I say thank you.

Abbreviations

aa	amino acid
aaRS	aminoacyl-tRNA synthetase
Ala	alanine
AMP	adenosine monophosphate
ATP	adenosine triphosphate
CTD	C-terminal domain
DTT	dithiothreitol
MALLS	multi-angle laser light scattering
MSC	multi-aminoacyl-tRNA synthetase complex
MST	microscale thermophoresis
PAGE	polyacrylamide gel electrophoresis
Pro	proline
ProRS	prolyl-tRNA synthetase
SEC	size exclusion chromatography
SLIM	site-directed ligase-independent mutagenesis
tRNA	transfer ribonucleic acid
WT	wild-type
ΔCTD	C-terminal domain truncation

Abstract

Aminoacyl-tRNA synthetases (aaRSs) catalyze bonds between tRNAs and cognate amino acids. The formation of cognate aminoacyl-tRNAs is critical for cell viability, as translating noncognate aa-tRNAs leads to amino acid misincorporation and alterations in protein structure and function. While aaRSs make errors in most organisms, proofreading mechanisms also exist; however, our understanding of this machinery is poorly understood, particularly outside of prokaryotes. Unlike *Escherichia coli (Ec)*, which possess a designated editing domain on prolyl-tRNA synthetase (ProRS), some bacteria like *Caulobacter crescentus* (*Cc*), multicellular eukaryotes like humans (*Homo sapiens*, *Hs*), and plants like *Arabidopsis thaliana* (*At*), encode a separate editing enzyme called ProXp-ala that hydrolyzes non-cognate Ala-tRNA^{Pro}. Sequence alignments have revealed that the homologs in plants all possess a C-terminal extension absent in other species. This C-terminal domain (CTD) has a conserved sequence with no known structure or function to date. This work seeks to elucidate the structure and function of this highly conserved CTD of *At* ProXp-ala. Bioinformatics analyses and homology modeling studies revealed that the plant-specific CTD may be an oligomerization motif. Full-length wild type (WT) and truncated C-terminal domain (ΔCTD) *At* ProXp-ala and *At* tRNA^{Pro} have been expressed and purified from *E. coli* and their storage conditions were optimized. Homology models predicted that the CTD plays a role in protein oligomerization, which was supported by size-exclusion chromatography with multiple angle laser light scattering experiments. *In vitro* kinetic deacylation assays demonstrated a 19-fold decrease in activity without the CTD and a positive correlation between activity and enzyme concentration for the ΔCTD variant. These data suggest that the CTD plays a significant role in the activity of *At* ProXp-ala and are consistent with a model in which truncation results in impaired binding of the ΔCTD mutant to tRNA^{Pro}. Preliminary microscale thermophoresis binding assays were performed to test this hypothesis but require further optimization. Comparisons of plant and human ProXp-ala activity revealed similar deacylation rates between the human enzyme and ΔCTD *At* ProXp-ala. If the *At* CTD does indeed improve tRNA binding, then appending this domain to human ProXp-ala is expected to similarly improve its substrate binding activity and future work will address this hypothesis. The emphasis of this work on translational fidelity in plants is a novel direction in aaRS research that will significantly expand our understanding of eukaryotic editing mechanisms. Our comparative study of plant and human ProXp-ala has implications as to the potential limitations of human translational machinery and may inform future studies on the mechanisms of human disease related to sub-optimal translational fidelity.

Chapter 1: Introduction

Translation is the process by which an organism's genetic information in the form of mRNA is used to construct functional proteins from amino acids at ribosomes within cells. Transfer RNAs (tRNAs) are ubiquitously expressed non-coding RNAs that are responsible for delivering amino acids to ribosomes. tRNAs are substrates for aminoacyl-tRNA synthetases (aaRSs), which create chemical bonds between tRNAs and amino acids through a process called aminoacylation or "charging"[1,2]. Aberrations in aminoacylation and translational fidelity have detrimental effects on cells, including causing degradation and apoptosis in mammalian cells, as well as contributing to neurodegeneration[3,4].

The aminoacylation reaction occurs in two steps: activation and transfer (**Fig. 1**). In the first step, an aaRS will bind an amino acid and use ATP to form an activated aminoacyl adenylate (aa-AMP) precursor, releasing pyrophosphate (PP_i). In the second step, the activated aa-AMP is transferred to either the 2' or 3' hydroxyl group (depending on the aaRS) at the 3' end of the bound tRNA (a region referred to as the "acceptor stem"), forming aminoacyl-tRNA (aa-tRNA)[5,6]. The aa-tRNA is then released from the aaRS active site and carried to the ribosome by elongation factors (EFs) for translation, leaving the synthetase available for another round of aminoacylation[5].

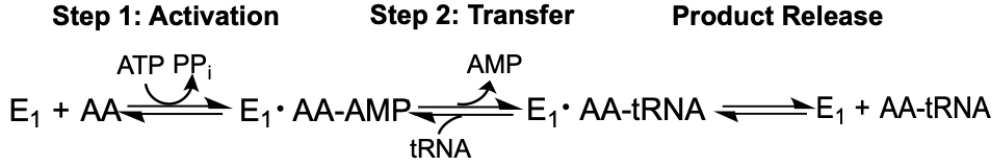


Figure 1. The aminoacylation reaction of aminoacyl-tRNA synthetases (E_1) occurs in two steps: activation of the amino acid (AA) by ATP, then transfer of the aminoacyl adenylate (AA-AMP) to the 3' acceptor stem of the tRNA. Dots separate bound species.

Most organisms possess at least twenty aaRSs, one for charging each of the common proteinogenic amino acids to all of its tRNA isoacceptors[5].The primary products of aminoacylation reactions are cognate aa-tRNAs, which are tRNAs that are correctly charged with the amino acid encoded by their respective anticodon loops. For example, prolyl-tRNA synthetases (ProRSs) generate cognate Pro-tRNA^{Pro} aminoacyl-tRNAs that are made of a proline amino acid linked to a tRNA with a proline-encoding anticodon (AGG, CGG, GGG, UGG in most organisms).

However, the process of aminoacylation is imperfect, and noncognate aa-tRNAs are often generated by aaRSs, which leads to mistranslation and the generation of mutant polypeptides, creating a necessity for translational proofreading machinery[7]. The large size of tRNAs provides ample elements for substrate discrimination, but the small sizes and similar structures of amino acids make amino acid selectivity a challenge for aaRSs[8,9]. Proteinogenic amino acids (and nonproteinogenic amino acids in appreciable concentrations) that are similar in molecular volume and composition to the cognate amino acid of a synthetase demonstrate the highest rates of mischarging (**Fig. 2**)[9]. Maintaining a steady intracellular pool of cognate aa-tRNAs is crucial for cells to be able to maintain fidelity during translation, since translation of mischarged aa-tRNAs introduces missense mutations into the nascent polypeptide, which has been found to result in severe structural and functional consequences such as misfolding of the protein product[1,4].

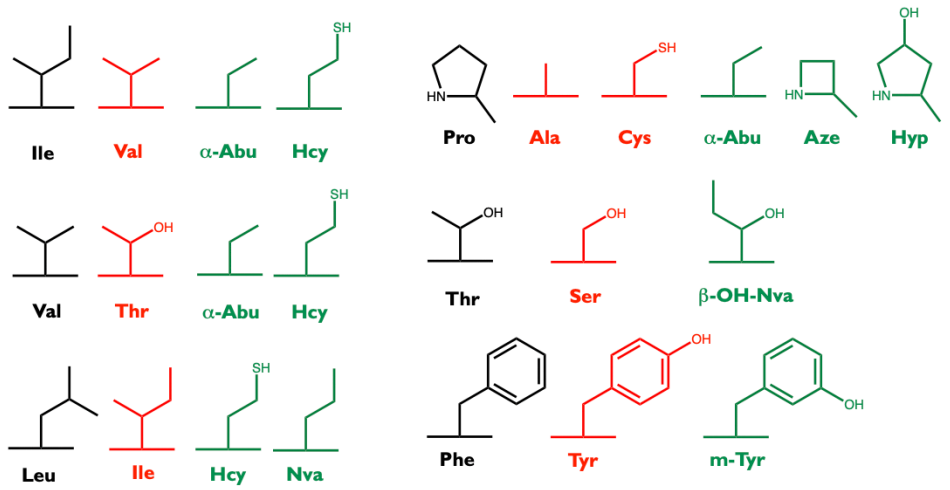


Figure 2. Amino acids frequently charged onto cognate tRNAs for selected synthetases. The cognate (black), proteinogenic (red) and nonproteinogenic (green) amino acids are shown. (Adapted from Jo Marie Bacusmo by Lexie Kuzmishin Nagy)

Selection for fidelity during translation is evident by the discovery of proofreading mechanisms that edit noncognate aa-tRNAs in all domains of life (**Fig. 3**)[10]. "Pre-transfer" editing occurs between the activation and transfer steps of aminoacylation (**Fig. 3**). For example, the *Ec* isoleucyl-tRNA synthetase (IleRS) recognizes a misactivated Val-AMP and hydrolyzes the bond before transfer of the amino acid to tRNA, successfully preventing mischarging[11-13]. However, it cannot always recognize Val-AMP at this step and may mischarge Val onto $tRNA^{Ile}$ producing noncognate $Val\text{-}tRNA^{Ile}$. To prevent mistranslation, IleRS employs post-transfer editing in *cis* by having the acceptor stem of bound mischarged $Val\text{-}tRNA^{Ile}$ interact with a separate editing domain on IleRS that will selectively remove Val from $tRNA^{Ile}$ in a deacylation reaction (**Fig. 3**). This form of editing is called "post-transfer" editing. The latter can also be performed after aa-tRNA product release in *trans*, either by the synthetase itself, or by a free-standing *trans*-editing deacylase (**Fig. 3**)[10,14].

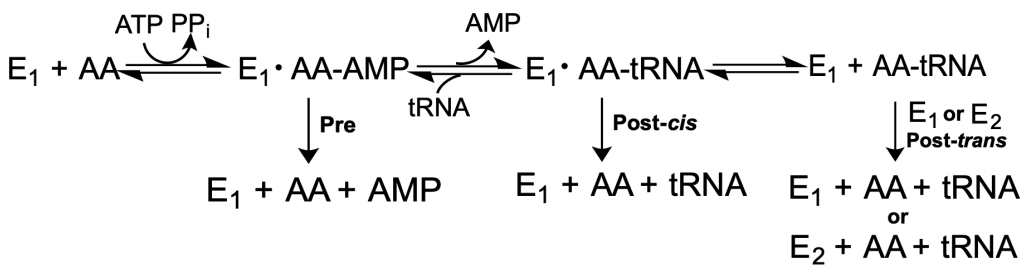


Figure 3. Translational fidelity is ensured through various editing mechanisms: pre-transfer hydrolysis of aminoacyl-adenylate (AA-AMP), *cis*-post-transfer deacylation of noncognate aminoacyl-tRNA (AA-tRNA) by an aminoacyl-tRNA synthetase (E_1) editing site, or *trans*-post-transfer deacylation after product release by an aminoacyl-tRNA synthetase (E_1) editing site or a free standing, *trans*-editing deacylase (E_2).

The exact combination of proofreading mechanisms that is employed to ensure a high concentration of a specific cognate aa-tRNA (e.g., Pro-tRNA^{Pro} or Ile-tRNA^{Ile}) varies between aaRSs and organisms, as there is variability in the concentrations of noncognate amino acids in different cellular contexts. The exact sequences and domain architectures of aaRSs also vary among species, even within the same domain of life. For example, *Ec* ProRS charges all four tRNA^{Pro} isoacceptors with Pro *in vivo*, and while most noncognate amino acids, like His are not subject to aminoacylation due to the steric constraints of the ProRS active site (**Fig. 4A**), *Ec* ProRS has been found to successfully activate and mischarge noncognate Ala and Cys onto tRNA^{Pro} *in vitro* (**Fig. 2**)[15,16]. While *Ec* ProRS exhibits some degree of pre-transfer editing similar to that of *Ec* IleRS, this activity is not sufficient to prevent accumulation of Ala-tRNA^{Pro}, making post-transfer editing necessary to fulfill the proofreading needs of *Ec* ProRS[17]. To clear Ala-tRNA^{Pro}, *Ec* ProRS contains an editing "insertion domain" (INS) that specifically deacylates mischarged Ala-tRNA^{Pro}, but not Cys-tRNA^{Pro}, though it is currently unclear whether the INS domain performs deacylation in *cis* or *trans* (**Fig. 4A)**[14,18]. To clear mischarged Cys-tRNA^{Pro}, *Ec* expresses

YbaK, a structurally homologous, INS-like, free-standing *trans*-editing protein that forms a ternary complex with ProRS and Cys-tRNA^{Pro}, then subsequently deacylates Cys-tRNA^{Pro} (**Fig. 4A**)[19-21]. This "triple-sieve" mechanism of active site steric hinderance, INS domain editing, and YbaK *trans*-editing ensures that Pro-tRNA^{Pro} is the only aa-tRNA^{Pro} available for translation in *Ec.*

A triple-sieve mechanism of Pro-tRNA^{Pro} proofreading is not employed by all species, and some bacteria even employ an alternative triple-sieve mechanism. For example, *Caulobacter crescentus* (*Cc*), exhibits Cys- and Ala-tRNA^{Pro} mischarging and expresses YbaK to clear Cys-tRNA^{Pro}, but lacks an INS domain on ProRS to clear mischarged Ala-tRNA^{Pro} (**Fig. 4B**)[22]. Instead, *Cc* encodes a free-standing ProXp-ala domain that is structurally homologous to the *Ec* INS domain and performs Ala-tRNA^{Pro} editing in *trans* (**Fig. 4B**)[23]. Together, the two *trans*-editing domains, ProXp-ala and YbaK, fulfill the tRNA^{Pro} proofreading needs of *Cc.*

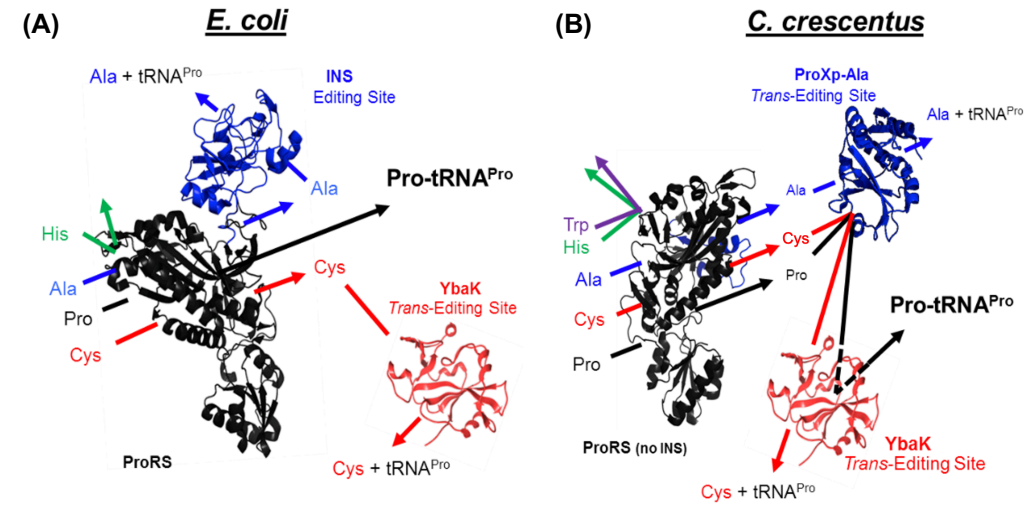


Figure 4. The "Triple Sieve" mechanism of **(A)** *Ec* ProRS employs steric effects (green), *cis*-post-transfer editing by the INS domain (blue), and *trans*-post-transfer editing by Ybak (red) to ensure that Pro-tRNA^{Pro} is the only aa- tRNA^{Pro} available for translation. Instead of an INS domain, **(B)** *Cc* ProRS encodes the *trans*-editing protein ProXp-ala to deacylate Ala-tRNA^{Pro}. (Adapted from Vargas-Rodriguez & Musier-Forsyth, 2013; Das et al., 2014)

Most of our understanding of translational fidelity mechanisms, especially with regard to free-standing *trans*-editing proteins, has come from research in prokaryotic systems. Recent bioinformatics analyses have revealed a large superfamily of nine INS-like proteins, including YbaK and ProXp-ala, many of which are largely uncharacterized (**Fig. 5A**)[14]. Of these proteins, most are found almost exclusively in prokaryotes; however, ProXp-ala is found in a number of eukaryotes, and it is the only known INS superfamily member widely expressed in animals to date (**Fig. 5B**)[14].

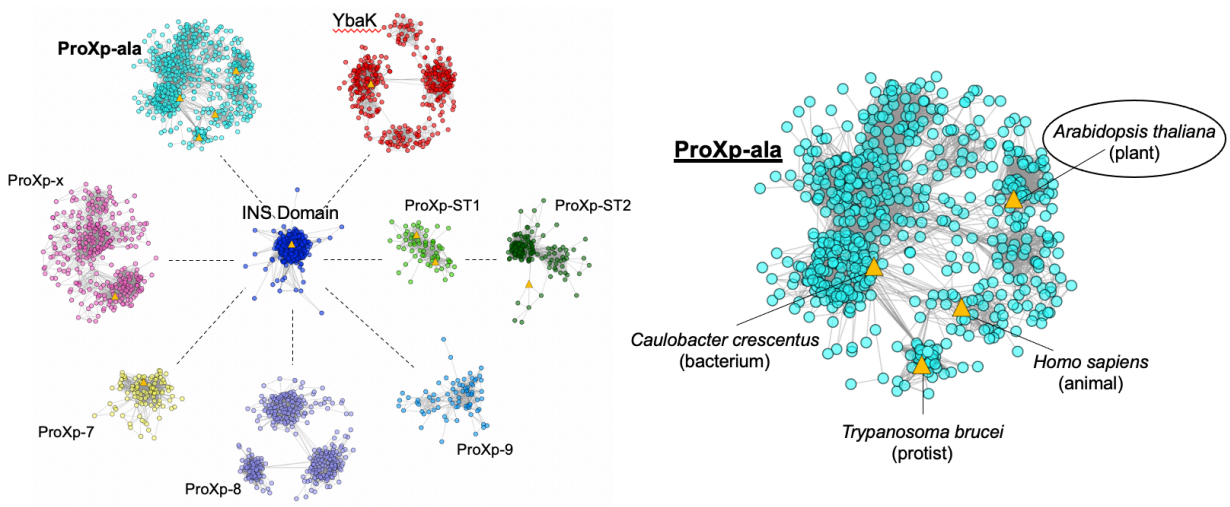


Figure 5. **(A)** Sequence similarity network analysis of the INS Superfamily of *trans*-editing domains generated and performed by Marina Bakhtina. Each node represents a collection of similar sequences. Length of the lines connecting nodes indicates degree of sequence similarity, revealing nine unique clusters of sequences. Six of these are consistent with established functional *trans*-editing factors and three are new clusters currently called ProXp-7, ProXp-8, and ProXp-9. **(B)** Magnified image of the ProXp-ala cluster revealing sub-clustering within kingdoms. Nodes containing ProXp-ala species investigated in the Musier-Forsyth lab are identified with triangles.

While *Cc* ProXp-ala has been well characterized both structurally and functionally, eukaryotic *trans*-editing mechanisms, particularly in plants, are poorly understood. Clustal Omega sequence alignments of ProXp-ala from bacteria, plants, and animals (**Fig. 6**), reveal that plant ProXp-ala contains an extension on the C-terminus that is approximately 140 residues in length. BLAST searches revealed that this extension shares no detectable similarity to any known protein. Previous models of aaRS evolution include monofunctional proteins that have developed more complex, multi-domain architectures in response to selective evolutionary pressure over time[24,25]. Thus, the presence of a conserved C-terminal domain (CTD) on a plant protein whose N-terminal domain is closely related to other known ProXp-ala enzymes in bacteria and animals suggests that this CTD may contribute a novel function to plant ProXp-ala and account for its retention throughout plant evolution. ProXp-ala sequences with C-terminal extensions have been found in all plant phyla, indicating that the evolution of the CTD occurred before the divergence of plants from other eukaryotes.

This thesis reports the insights gained in understanding the structure, function, and evolution of ProXp-ala in the model plant system *Arabidopsis thaliana* via bioinformatics and computational modeling (**Chapter 2**), and functional comparisons of full-length wild type (WT) and genetically truncated C-terminal domain (ΔCTD) *At* ProXp-ala (**Chapters 3 and** 4). Translational fidelity in plants is a previously unexplored direction in aaRS research, thus this work will significantly expand our understanding of eukaryotic editing mechanisms, especially in plants.

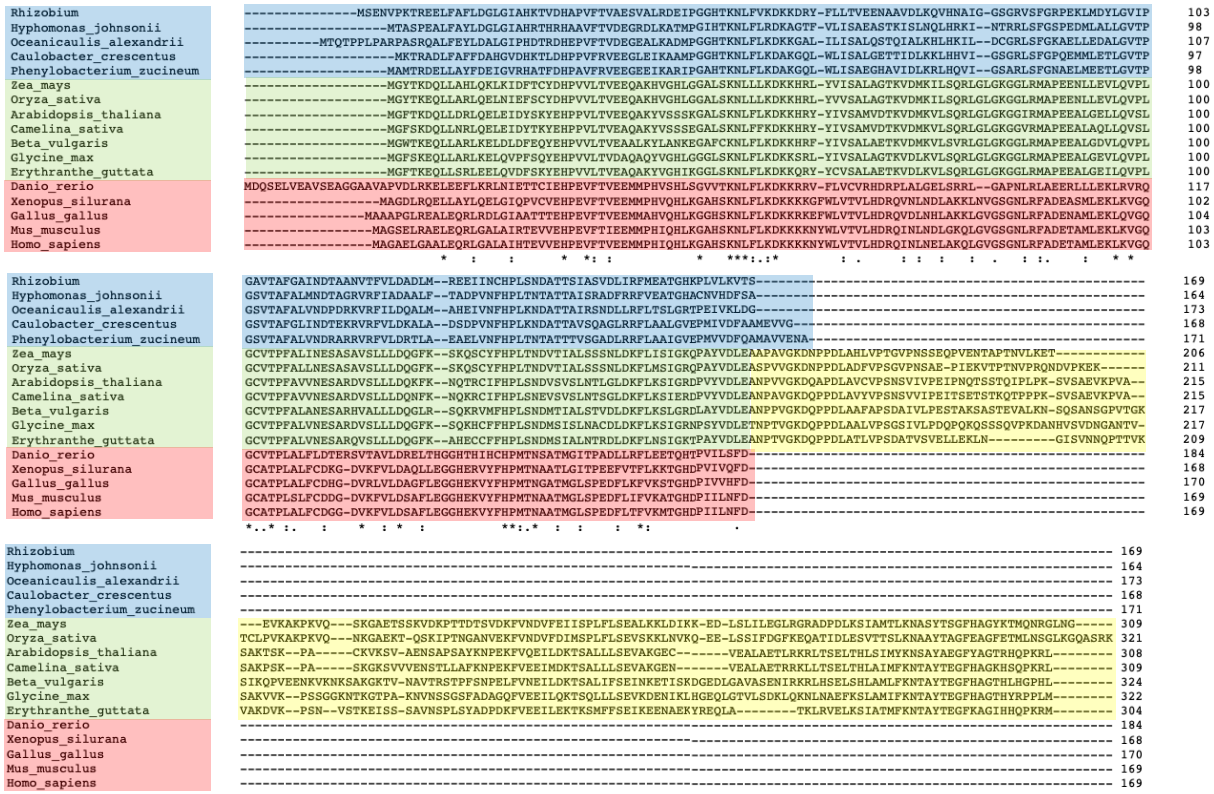


Figure 6. Alignment of ProXp-ala sequences from bacterial (top, blue), plant (middle, green), and animal (bottom, red) species. The unique plant CTD is highlighted in yellow.

.

Chapter 2: Bioinformatics and Structural Predictions of *Arabidopsis thaliana* ProXp-ala

Introduction

While the structures of *Ec* Ybak (PDB ID: 2DXA) and *Cc* ProXp-ala (PDB ID: 5VXB), have been determined previously by X-ray crystallography, the structure of *At* ProXp-ala, including its unique C-terminal domain, is unknown. Initial alignments comparing *Cc*, *Hs*, and *At* ProXp-ala confirm the notable homology among approximately the first 165 N-terminal amino acids, after which *Cc* and *Hs* reach their C-terminus and *At* continues with a stretch of approximately 145 residues (**Fig. 7**). This pattern of homology suggested that the N-terminal 165 amino acids of *At* ProXp-ala likely comprise a catalytic core that has preserved the Ala-tRNA^{Pro} editing activity as observed in *Cc* and *Hs* ProXp-ala. However, the addition of a conserved CTD in plants (**Fig. 6**) suggests that plant ProXp-ala may have evolved enhanced or novel functions that have been retained throughout evolution. This chapter will discuss some of the *in silico* techniques and software programs that have been employed to predict the secondary, tertiary, and potentially quaternary structures of *At* ProXp-ala, which informed the types of functions that were probed in later *in vitro* studies.

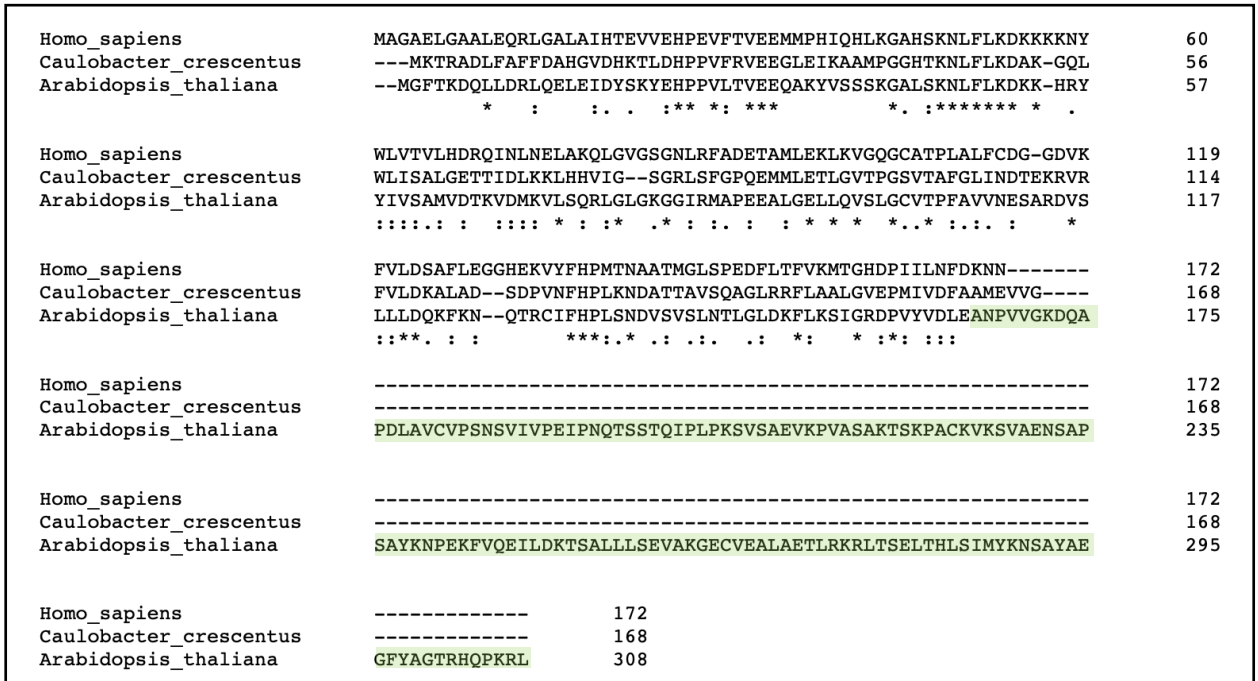


Figure 7. The *At* ProXp-ala sequence exhibits notable conservation with *Cc* and *Hs* ProXp-ala throughout the first 165 residues of the protein. Residues 166-308 (green) make up a C-terminal extension with no recorded structure or function.

Materials and Methods

Porter 4.0, PaleAle 4.0

The amino acid sequence of *At* ProXp-ala was used as a query for the Porter 4.0, PaleAle 4.0 protein structure prediction service. The Porter program outputs a secondary structure prediction for each residue (helix, strand, or coil) and the PaleAle program predicts the relative solvent accessibility (very buried, somewhat buried, somewhat exposed, very exposed) for each residue as well[26]. Together, these data can be used to predict the domain architecture and overall structure of a query polypeptide sequence.

SWISS-MODEL

The amino acid sequence of *At* ProXp-ala was used as a query for the SWISS-MODEL protein structure homology-modeling server. The query returned a number of INS-like domain proteins, including the *Cc* ProXp-ala crystal structure (5VXB), which was used to generate a homology model for *At* ProXp-ala, as well as *Hs* ProXp-ala which similarly lacks a known structure. SWISS-MODEL is only able to generate models that contain stretches with similar identity between the template and query sequence. Thus, the software was only able to generate a model using the N-terminal core of *At* ProXp-ala since there was not enough sequence identity to the CTD. To gain insights into the potential tertiary structure of the *At* ProXp-ala CTD, a new search was performed using only the C-terminal residues (164-308) of the protein. While homology models are not a sufficient substitute for actual structural data, probing the sequence of the CTD for homology to known structures provided insight into the possible functions that this extension could exhibit *in vitro* and *in vivo,* as discussed below.

PyMOL Molecular Modeling Software

This program visualized the .pdb files generated by SWISS-MODEL. The homology models of *Hs* and *At* ProXp-ala were overlaid on top of the template structure of *Cc* ProXp-ala (5VXB) and very strong structural similarity was observed, as nearly all secondary structures overlapped perfectly among the three proteins. To confirm that structural similarities were not merely the result of the program forcing the query sequence to fit the template model, 5VXB was also used as a template for the CTD sequence of *At* ProXp-ala; these sequences have no notable homology to one another. The resulting structure lacked the significant secondary structure overlap seen in the other model and was almost entirely composed of random coils (data not shown).

Results and Discussion

Predicted Domain Architecture of At *ProXp-ala*

The Porter,PaleAle program output revealed important structural predictions for the C-terminal extension of *At* ProXp-ala (**Fig. 8**). The N-terminal catalytic core of the protein exhibits a mixture of secondary structures and random coils, with significant variation in solvent accessibility. These observations are consistent with the results expected for a soluble, globular single-domain protein, and are also consistent with the results found when using *Cc* or *Hs* ProXp-ala as query sequences (data not shown). However, the C-terminal extension appears to lack this globular structural signature, and instead is predicted to contain a disordered, highly exposed random coil that is followed by a long α-helical domain. These findings suggest that the C-terminal extension contains a potentially functional, long α-helical nucleic acid or protein interaction domain that is connected to the N-terminal catalytic core via a long, random coil linker region

Homology-Based Functional Predictions of At ProXp-ala Domains

The SWISS-MODEL program and PyMOL imaging software were used to generate and visualize homology models of *Hs* ProXp-ala and the N-terminal core of *At* ProXp-ala. Their structures can be seen overlaying the *Cc* ProXp-ala template structure (5VXB) from which they were generated (**Fig. 9**). There is clear overlap between all three structures and given the similar primary sequence and secondary structure predictions of the three proteins, this predicted structural conservation at the tertiary level is consistent with the hypothesis that *At* ProXp-ala retains the Ala-tRNA^{Pro} editing function exhibited by *Cc* and *Hs* ProXp-ala.

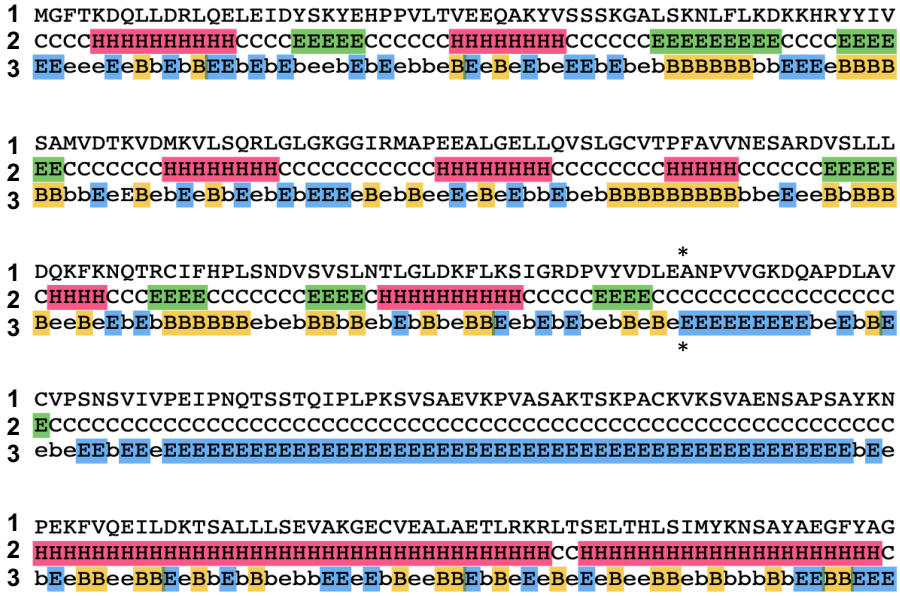


Figure 8. Porter,PaleAle structural predictions of *At* ProXp-ala. **1**: *At* ProXp-ala sequence. **2**: Porter secondary structure prediction, H=helix (red), E=β-strand (green), C=coil (uncolored). **3**: PaleAle relative solvent accessibility prediction, B=very buried (yellow), b=somewhat buried (uncolored), e=somewhat exposed (uncolored), E=very exposed (blue). * marks the apparent separation between the N-terminal catalytic core and the C-terminal extension.

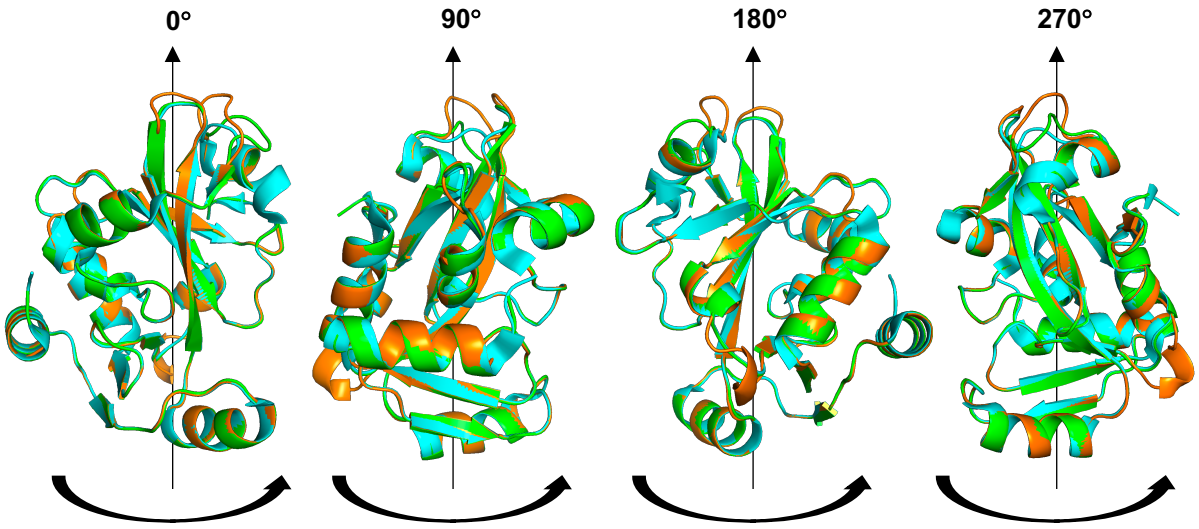


Figure 9. Homology models of *At* (green) and *Hs* (orange) ProXp-ala generated by SWISS-MODEL overlaying the *Cc* (blue) ProXp-ala template structure 5VXB. Each image is an approximate 90 degree rotation along a vertical axis relative to the image to its left.

Since the initial homology model in **Fig. 9** could not be generated for the CTD, the sequence of C-terminal residues was used as a SWISS-MODEL query to probe for homologous sequences and structures for just the CTD. Initial searches returned a template structure with 21.05% sequence identity to latter half of the CTD including a highly helical region predicted previously in **Fig. 8**. This protein, *Chlamydia trachomatis* (*Ct*) plasmid gene product 3 (Pgp3), is a large, three-domain protein that was crystallized as a homotrimer. An individual image of the CTD homology model can be seen in **Fig. 10B**, and an overlay of the *At* ProXp-ala CTD on the *Ct* Pgp3 trimer is shown in **Fig. 10C**.

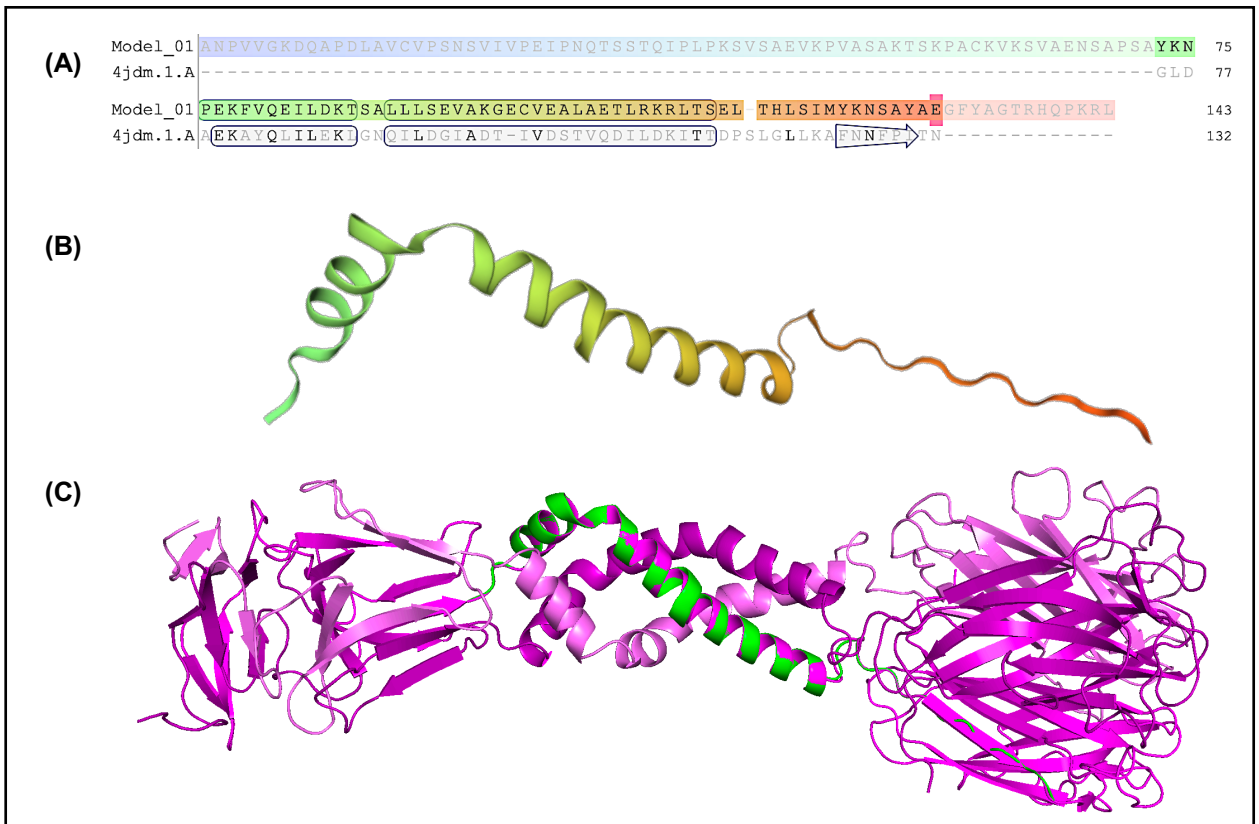


Figure 10. Homology modeling of the *At* ProXp-ala CTD by SWISS-MODEL using the *Ct* Pgp3 crystal structure template. **(A)** Sequence alignment of the region with 21% identity between the *At* ProXp-ala CTD (Model_01) and *Ct* Pgp3 (4jdm.1.A). **(B)** Predicted structure of helical region of *At* ProXp-ala CTD. **(C)** Overlay of *At* ProXp-ala CTD homology model (green) on one of three identical chains of the *Ct* Pgp3 homotrimer (purple).

Based on the aforementioned secondary and tertiary structural predictions, the CTD of At ProXp-ala is hypothesized to contain a long linker region that connects the N-terminal catalytic deacylase core to an α-helical domain. The homology of the coiled region to a *Ct* Pgp3 homo-trimerization domain suggests that this helix may facilitate protein-protein interactions, possibly even homo-oligomerization, that may be critical for performing both its canonical Ala-tRNA^{Pro} editing function, as well as any novel noncanonical functions *in vivo*. Proteolysis experiments to investigate this predicted domain architecture model are planned for the near future.

Acknowledgements

I would like to thank Jun-Kyu Byun for providing me with the ProXp-ala sequences used as queries in these studies and Dr. William Cantara for his suggestions on which programs and software tools would best support the goals of this project.

Chapter 3: Purification of *Arabidopsis thaliana* ProXp-ala and $tRNA^{Pro}$

Introduction

The results of the bioinformatics and modeling studies described in **Chapter 2** have suggested that *At* ProXp-ala is very likely to possess Ala-$tRNA^{Pro}$ deacylation activity catalyzed by its N-terminal catalytic core, while homology modeling of the C-terminal extension suggests that it may be a protein-protein interaction domain. To probe these hypothesized functions and any other potential functions of the C-terminal extension of *At* ProXp-ala, the relevant macromolecules needed to be purified.

Full-length (wild-type, WT) *At* ProXp-ala and a truncated variant containing a C-terminal domain deletion (ΔCTD) were prepared. WT *At* ProXp-ala is a 35-kDa protein that is 308 residues in length, averaging around 180% of the size of ProXp-ala found in *Cc* and *Hs*, which are 168 and 172 residues long, respectively (**Fig. 7**). To generate an effective ΔCTD mutant, all of the critically conserved residues necessary for enzymatic function in ProXp-ala expressed in other species were retained in the mutant. The truncation was designed and constructed by Jun-Kyu Byun. ΔCTD is a 20-kDa protein lacking residues 166-307, which are predicted to form a disordered random coil linker and a long α-helical domain (**Fig. 8**). Both constructs were designed with N-terminal His6 tags with sizes of approximately 2-kDa, resulting in molecular weights of 37-kDa and 22-kDa for WT and ΔCTD, respectively.

In addition to purifying the WT and mutant proteins, an appropriate *At* $tRNA^{Pro}$ substrate was necessary for the functional assays. Like most eukaryotes, *At* encodes a number of $tRNA^{Pro}$ isoacceptors, each of which is 76 nucleotides in length with a CCA-3' acceptor stem and one of the four proline anticodons at nucleotide positions 34-36: UGG, GGG, AGG, or CGG. For this

study, the sequence of the UGG anticodon was chosen because 70% of the tRNAPro encoding genes in *At* use this anticodon according to the Genomic tRNA Database (gtRNAdb).

Materials and Methods

Electrocompetent E. coli Transformation with Plasmids

To purify WT *At* ProXp-ala, the DNA sequence of Kyoto Encyclopedia of Genes and Genomes (KEGG) entry AT1G44835 with an N-terminal His-tag was purchased by Lin Chen and cloned by her into a pET15b vector, which contains an upstream inducible lac operon and an ampicillin resistance (ampR) cassette. Electrocompetent BL21(DE3)RIL *E. coli* (chloramphenicol resistant, camR) were transformed with this plasmid via electroporation and grown overnight at 37°C after plating on solid LB media with ampicillin (amp) and chloramphenicol (cam). A single colony was used to inoculate a 35 mL starter culture of liquid LB+amp+cam media overnight. The next day, the starter culture was added to 1 L of LB+amp+cam liquid media, and once the culture was grown to an OD_{600} of 0.6, a 1 mL aliquot was withdrawn and used to prepare a WT *At* ProXp-ala pET15b BL21(DE3)RIL glycerol stock (ampR, camR), which was stored at -80°C. The remaining culture was used for protein purification as outlined below.

A CTD deleted *At* ProXp-ala (ΔCTD) construct was generated by Jun-Kyu Byun via site-directed, ligase-independent mutagenesis (SLIM)[27]. BL21(DE3)RIL *E. coli* were transformed with this plasmid and used to prepare a ΔCTD glycerol stock as described for the WT sequence.

The desired tRNA construct containing the *At* $tRNA^{Pro(UGG)}$ sequence with an upstream T7 promoter flanked by BstNI endonuclease restriction enzyme recognition sites was designed and purchased by Lin Chen and cloned by her into a pUC57 vector (ampR). This plasmid was transformed into electrocompetent XL1-Blue *E. coli* via electroporation and grown on solid

LB+amp media. A single colony was selected and used to inoculate a 100 mL LB+amp liquid media culture and grown to an OD_{600} of 0.6. A 1 mL aliquot was withdrawn and used to prepare a WT *At* $tRNA^{Pro(UGG)}$ pUC57 XL1-Blue glycerol stock, which was stored at -80ºC.

Protein Purification

To purify WT and ΔCTD *At* ProXp-ala, the appropriate pET15b BL21(DE3)RIL glycerol stock was streaked onto LB+amp+cam plates. A single colony for each construct was selected and grown in a 35 mL starter culture of liquid LB+amp+cam media at 37°C overnight. The next day, the starter culture was used to inoculate a 1 L culture of liquid LB+amp+cam media and allowed to grow for 2-3 h until the OD_{600} reached 0.6. At this point, the culture was placed on ice for 20 min to induce expression of cold-shock chaperone proteins which enhance protein solubility. Isopropyl β-D-1-thiogalactopyranoside (IPTG), an allolactose molecular mimic, was added to the media (100 μM final concentration), triggering the upstream lac operon to induce transcription of the WT or ΔCTD *At* ProXp-ala gene. Induction was performed overnight at 25°C and was confirmed via denaturing sodium dodecyl sulfate 12% (w/v) polyacrylamide gel electrophoresis (SDS-PAGE) (**Fig. 11A**).

The induced cultures were grown overnight, and the cells were harvested the next day by centrifugation (6,000 relative centrifugal force (rcf), 15 min, 4°C). The total cell pellet mass was determined, and one cOmplete, Mini Protease Inhibitor Cocktail tablet (Roche) was added for each gram of cell pellet. The cells and tablets were mixed in 10 mL/gram of pellet of lysis buffer (50 mM NaP_i pH 7.5, 600 mM NaCl, 20 mM β-mercaptoethanol (β-ME, a reducing agent), 10% v/v glycerol) with 100 mg/gram of pellet of lysozyme from chicken egg white (Sigma-Aldrich). Once the cells were fully resuspended and the tablets were dissolved, the solution was incubated on ice

for 30 min. Following incubation, the cells were lysed further via sonication on ice (9 rounds of 40 s of 1 s on/off pulses followed by 40 s of rest). The lysed cell fragments were centrifuged (27,000 rcf, 30 min, 4°C), and the supernatant was removed with a syringe and filtered through 0.45 μm syringe filters.

Ni-NTA affinity columns were prepared at 4°C by packing 5 mL HIS-Select Ni-NTA Resin (Sigma-Aldrich) between the frits of 5 mL disposable polypropylene columns (QIAGEN). The columns were washed with 50 mL of pure water and equilibrated with 20 mL of the aforementioned lysis buffer. The filtered lysates for each protein were loaded onto separate columns, and the flow-throughs were collected. A 9-step gradient from 0 to 250 mM (0, 5, 10, 20, 30, 40, 60, 80, and 250 mM) imidazole, a His-side chain mimic, was used to elute the proteins and the fractions were collected separately before being run on a 12% SDS-PAGE gel to locate the fractions with the purest protein (**Fig. 11B-C).**

If the elution fractions were of sufficient purity, Amicon Ultra-15 Centrifugal Filter units (WT: 30,000 MWCO; ΔCTD: 10,000 MWCO) were used to concentrate the desired fractions by centrifugation (4,000 rcf, 10 min, 4°C). After the sample volume was reduced to below 1500 μL, the filter unit was filled with the appropriate protein storage buffer (WT: 500 mM NaP_i pH 6.0, 500 mM NaCl, 1 mM DTT; ΔCTD: 500 mM NaP_i pH 7.5, 150 mM NaCl, 1 mM DTT) and centrifugation was repeated three times to ensure sufficient buffer exchange. After the third round of exchange, the sample volume was reduced to 600 μL and combined in a 1:1 ratio with 80% (v/v) glycerol for storage at -20°C.

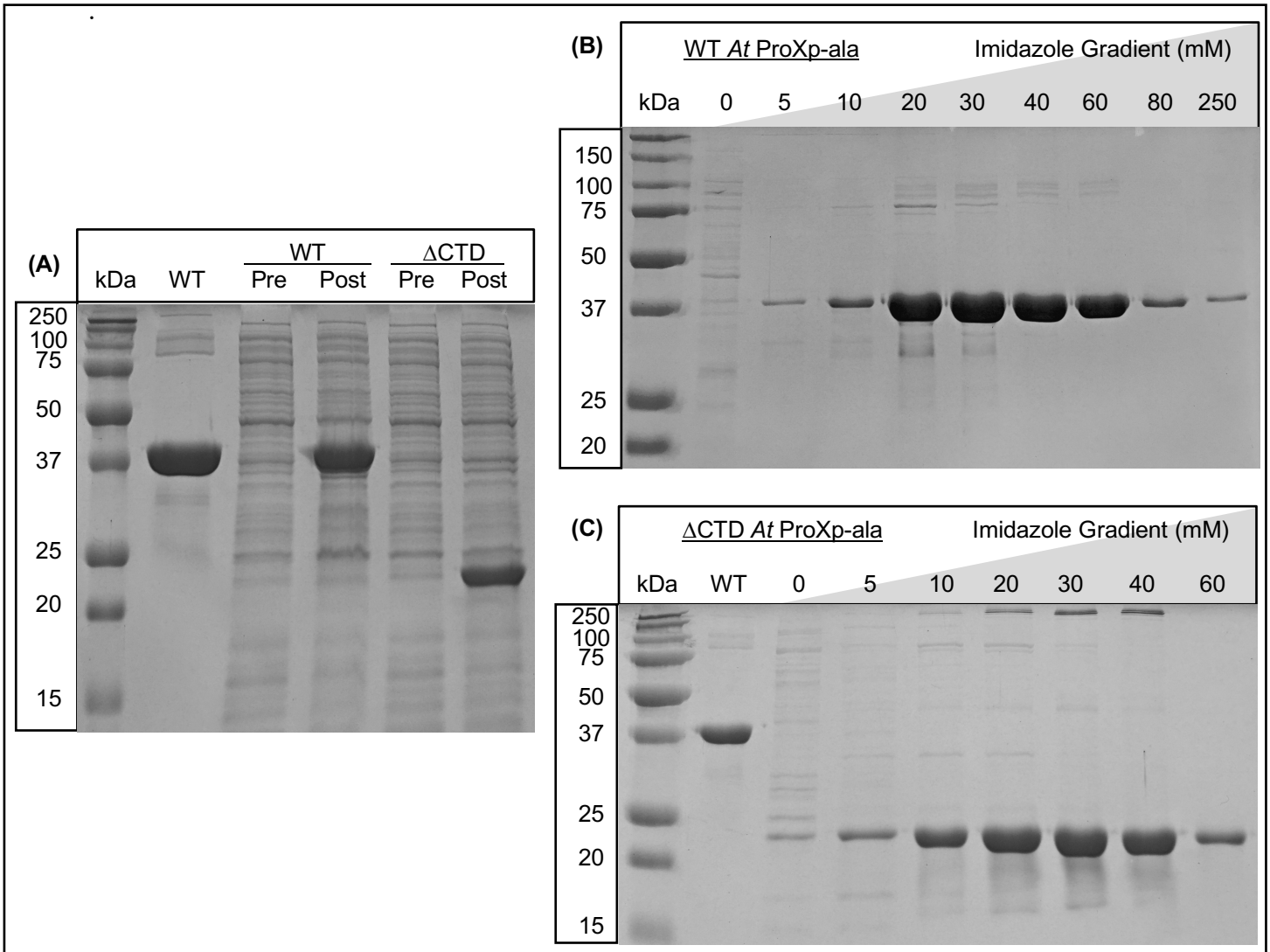


Figure 11. SDS-PAGE gels showing protein induction and purification results. Gels contained 12% polyacrylamide and were electrophoresed at 200 V for 45 min. **(A)** Induction test: 1 mL aliquots were drawn from each cell culture before (Pre) and after (Post) addition of IPTG. Each aliquot was centrifuged (17,000 rcf, 5 min, 4 °C) and the resulting pellet was resuspended in a 1:1 ratio of 2X protein loading buffer. Previously purified WT protein is used as an additional reference marker. Ni-NTA affinity column purification showing eluted fractions for WT **(B)** and ΔCTD **(C)** *At* ProXp-ala. Lanes contain 10 μL of the collected eluant after sequential loading of 20 mL (0-5 mM), 10 mL (10 mM), or 5 mL (20-250 mM) of lysis buffer with the indicated concentration of imidazole. WT is used as an additional reference marker in **(C)**.

If high and low molecular weight impurities were present after the elution step, purification was continued with size-exclusion chromatography (SEC). The desired fractions were combined in Amicon Ultra-15 Centrifugal Filter units (WT: 30,000 MWCO; ΔCTD: 10,000 MWCO) and centrifuged (4,000 rcf, 10 mins, 4°C) to concentrate the sample to less than 1 mL. The sample was then injected onto a Superdex 75 10/300 gel filtration column (Cytiva) equilibrated with the appropriate protein storage buffer (see above) using a Bio-Rad fast protein liquid chromatography (FPLC) instrument. After sample injection, 240 mL (two column volumes) of the desired storage buffer was pumped through the column by the instrument at a rate of 0.25 mL/min. Fractions (2 mL) were collected throughout loading and elution, and a UV laser was used to detect the presence of protein in each fraction. The resulting UV chromatogram was used to identify fractions containing protein. Aliquots from these fractions were electrophoresed on 12% SDS-PAGE gels, and fractions containing the desired protein were combined, concentrated, and buffer exchanged into protein storage buffer in Amicon Ultra-15 Centrifugal Filter units, as described previously. Proteins were combined 1:1 with 80% (v/v) glycerol and stored at -20°C.

T7 RNA Polymerase in vitro *tRNA Transcription and Gel Purification*

The previously described *At* $tRNA^{Pro(UGG)}$ pUC57 XL1-Blue glycerol stock was streaked onto LB+amp plates. A single colony was selected and used to inoculate a 35 mL starter culture of liquid LB+amp media overnight. The next day, the starter culture was added to a 1 L culture of liquid LB+amp media and grown at 37°C for 2-3 h until OD_{600} was between 0.4 and 0.6. The cells were then harvested by centrifugation, and the DNA was extracted following the QIAGEN Maxi Prep protocol. The concentration of collected DNA was measured using UV-Vis spectroscopy on a NanoDropOne instrument.

To generate a stable template for transcription, 100 μg of plasmid DNA was digested with 15 μL of 1U per 1 μg/μL BstNI endonuclease in 200 μL 1X NEB Buffer 3.1. A 5 μL sample of the digestion reaction was electrophoresed on a 1% (w/v) agarose gel containing ethidium bromide at 100 V for 60 min before it was imaged with UV light to confirm the presence of the desired 215 base pair band determined by plasmid mapping (**Fig. 12A**). A 1 mL *in vitro* transcription of the 100 μg of BstNI digested DNA product using 50 μL of homemade T7 RNA polymerase (volume determined based on assay of enzyme efficiency) was performed at 37°C for 3.5 h in 1X NEB T7 transcription buffer, 4 mM NTPs, 20 mM $MgCl_2$, 5 mM DTT, and 5 μM PP_i-ase[28]. The transcription product was then electrophoresed on a denaturing 12% (w/v) polyacrylamide/7 M urea gel at 450 V for 3 h. The gel was imaged under a UV light and the tRNA band was excised from the gel, crushed, and soaked in 15 mL of RNA extraction buffer (500 mM NH_4OAc, 1 mM EDTA, pH 8.0) overnight with vigorous shaking at 37°C (**Fig. 12B**). After extraction, the $tRNA^{Pro(UGG)}$ solution was centrifuged (3000 rcf, 5 min, 4°C) and the supernatant was filtered through 0.45 μm filters to prevent gel pieces from continuing through the purification.

The filtered tRNA was concentrated to a volume of 400 μL via butanol extraction and then precipitated via overnight storage at -80°C in 800 μL of cold 100% ethanol and 40 μL of 3 M sodium acetate (pH 5.0). The next day, the samples were centrifuged (13,300 rcf, 45 min, 4°C) and the ethanol buffer was decanted leaving behind a white tRNA pellet. The pellet was washed in 300 μL of 70% (v/v) ethanol, centrifuged (13,300 rcf, 20 min, 4°C), decanted again, and finally left to air dry for an hour. The dried pellet was then resuspended in sterilized, filtered, MilliQ water, and stored at -20°C. The concentration of the tRNA stock was determined by NanoDropOne UV-Vis absorbance at 260 nm.

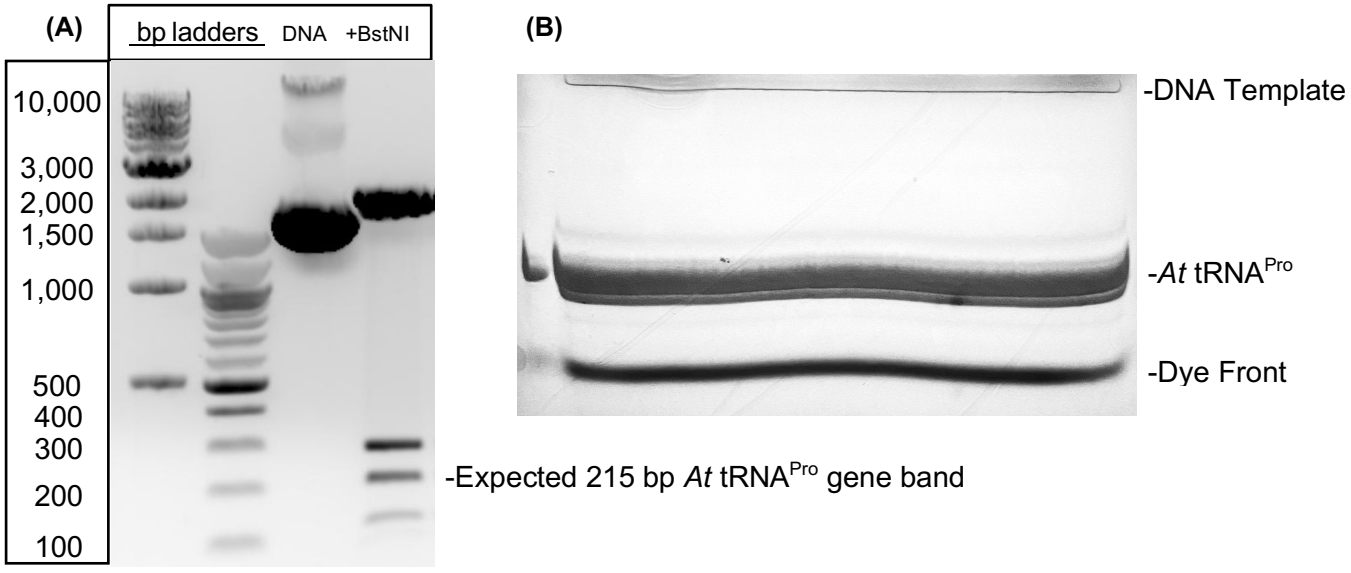


Figure 12. *At* $tRNA^{Pro}$ purification gels showing successful plasmid digestion and *in vitro* transcription. **(A)** QIAGEN Maxi-Prep purified DNA analyzed by 1% (w/v) agarose gel electrophoresis (100 V for 60 min) before and after addition of BstNI endonuclease. The desired $tRNA^{Pro}$ gene with a T7 promoter is expected to migrate as a 215 bp fragment. **(B)** T7 RNA polymerase-mediated *in vitro* transcription product analyzed by 12% (w/v) polyacrylamide/7 M urea gel electrophoresis (450 V for 3 h). Previously purified *At* $tRNA^{Pro}$ was run on the left side as a control. Only the largely enriched band was excised for subsequent purification.

Bradford Assay for Protein Concentration

Since the sequence of *At* ProXp-ala lacks tryptophan (Trp) residues necessary for determining protein concentration using A_{280}, the Bradford assay was used to determine the concentration of the purified WT and ΔCTD *At* ProXp-ala protein stocks[29]. In the presence of protein, Bio-Rad Bradford Dye changes from a reddish-brown color to a blue species that absorbs at 550 nm. This assay employs a known standardized protein solution to proportionately determine the concentration of the protein of interest. Bovine serum albumin (BSA) that was dissolved from powder to 10 mg/mL and diluted 10-fold to 1 mg/mL (verified using ε_{280} of 0.67 L Da^{-1} cm^{-1} for BSA) was used as the standard.

Standard protein solutions (800 μL) were prepared using 0 to 14 μL of 1 mg/mL BSA titrated in 2 μL increments. Bio-Rad Bradford Assay Dye (200 μL) was added to each solution-containing cuvette and the A_{550} was measured via NanoDropOne after 15 min. The A_{550} of each sample was plotted against the volume of BSA added and the fitted linear change in A_{550} was used to calibrate the slope of the BSA line.

Using the proteins of interest (WT and ΄ΔCTD *At* ProXp-ala), 800 μL protein solutions were similarly prepared, ranging from 2-10 μL of protein stock solution in 2 μL increments. The A_{550} was measured 15 min after the addition of 200 μL Bio-Rad Bradford Assay Dye, and the measured A_{550} readings were plotted against solution protein volume and fit to a line. Using the ratio of the slope of the normalized BSA line and the known ~1 mg/mL BSA concentration as a proportionality factor, the fitted slope of the line for the proteins of interest could be used to determine the concentration of target protein.

Buffer Stability Assay

It was observed via SDS-PAGE that months after storage in the original protein storage buffer used for previously studied ProXp-ala species like *Cc* and *Hs* (500 mM NaP_i pH 7.5, 150 mM NaCl, 1 mM DTT), WT ProXp-ala stocks would exhibit increasingly larger fractions of the total protein in high MW bands relative to the monomeric 37-kDa band that predominates after purification (data not shown). This observation suggested that these buffer conditions were promoting formation of high MW oligomers or aggregates that may influence functional protein activity. To optimize the storage conditions of the WT protein, 50 μL of highly pure protein (~100 μM) were buffer exchanged in triplicate into 18 different NaP_i buffers of different pH (6.0, 7.0, 8.0), salinity (NaCl: 0 mM, 250 mM, 500 mM), and reducing environment (DTT: 0 mM, 5 mM).

A single replicate of each condition was stored in 40% (v/v) glycerol at -80°C, -20°C or 4°C. The buffer that retained the greatest monomer to oligomer ratio upon visual inspection was ultimately selected for improved long-term storage of this protein.

Circular Dichroism Spectroscopy

Since the CTD is predicted to be highly α-helical, truncation of this domain would be expected to produce a large reduction in the amount of α-helix relative to WT. To measure the difference in secondary structure composition between WT and ΔCTD *At* ProXp-ala, Jun-Kyu Byun performed circular dichroism (CD) spectroscopy, which measures changes in molar ellipticity (θ) at various wavelengths. A strong dip in molar ellipticity at 208 and 220 nm corresponds to the characteristic signal of α-helices. Significant differences in the sizes of these peaks indicate differences in the α-helix composition of the proteins of interest.

Size Exclusion Chromatography

The homology of the CTD of *At* ProXp-ala to a trimerization domain of *Ct* Pgp3 suggested that the CTD may function as an oligomerization domain. To determine the oligomeric states sampled by WT and ΔCTD *At* ProXp-ala, purified WT or ΔCTD was loaded by Dr. William Cantara onto a size exclusion chromatography column that flowed through a multi-angle laser light scattering detector. This instrument then used UV absorbance at 280 nm and the retention volume of the sample to generate a spectrum. The column was calibrated so that the retention volume of a peak could be used to calculate the molecular weight of the oligomeric species represented by that peak.

Results and Discussion

Protein and tRNA Purification Yields and Purity

Induction of both WT and ΔCTD *At* ProXp-ala was consistently successful, and both proteins remained primarily in the soluble fraction during the purification process as shown in the representative gel in **Fig. 11A**. Both proteins eluted from the Ni-NTA affinity column, although there were often high molecular weight contaminants observed during WT purification. Passing this protein through a SEC column did not significantly improve purity and the intensity of the high molecular weight bands increased consistently over time (data not shown).

These results suggested that WT *At* ProXp-ala is a highly soluble protein that may form a single dominant oligomeric state in solution but oligomerizes over time. We hypothesized that the high molecular weight bands may be tightly interacting oligomers that are not dissociated even with denaturing conditions (SDS-PAGE and boiling water bath for 10 min). Further support for this hypothesis comes from the fact that the lowest high-molecular weight band exhibits a molecular weight that would be consistent with a dimer (~80 kDa), suggesting that these bands are likely due to irreversible higher order structures and not random protein aggregation (**Fig. 13**).

After gel purification, approximately 200 μg of tRNA was recovered for each 1 mL T7 RNA polymerase transcription reaction from 100 μg of original DNA template. The tRNA ran consistently on the urea-PAGE gels and could be excised for extraction from the gel (**Fig. 12B**).

ProXp-ala Buffer Storage Optimization

When comparing buffer storage conditions after two weeks, the samples stored at -20°C were indistinguishable from one another, confirming prior observations that this potential oligomerization occurs over long periods of time during -20°C storage. However, marked

differences were observed between the protein stocks stored at 4°C (**Fig. 13A-B**). pH had the greatest effect on stability, as all of the stocks stored at pH 6.0 (**Fig. 13A**, lanes 2-4) showed less high MW banding than those at pH 7.0 (**Fig. 13A**, lanes 5-7), which showed less than those stored at pH 8.0 (**Fig. 13A**, lanes 8-10). Additionally, salinity also had an effect on stability; for each pH tested, those stored in high salt conditions (**Fig. 13A**, lanes 2/5/8) resulted in fewer high MW bands compared to those stored with less salt (**Fig. 13A**, lanes 3/6/9) or no salt (**Fig. 13A**, lanes 4/7/10). Lastly, by comparing the lanes in **Fig. 13A** to those in **Fig 13B**, which only differ by the presence or absence of 5 mM DTT, respectively, the reducing environment also appeared to have a slight effect at reducing the formation of high MW bands (i.e. **Fig. 13A**, lane 5 and **Fig. 13B**, lane 4).

The best storage conditions found for the WT protein were low pH, high salt, and high DTT (**Fig. 13A, lane 2**). Following these results, the WT protein was henceforth stored in 500 mM NaP_i pH 6.0, 500 mM NaCl, 1 mM DTT and 40% glycerol. Notably fewer high MW species have been observed over time since this buffer was adopted. While this optimization assay was not performed for the ΔCTD based on its stability in the previously established buffer conditions, the fact that the ΔCTD protein does not exhibit the same level of aggregation implies that the CTD likely has a role in this activity and could potentially be serving as a homo-oligomerization interaction domain (**Fig. 12**).

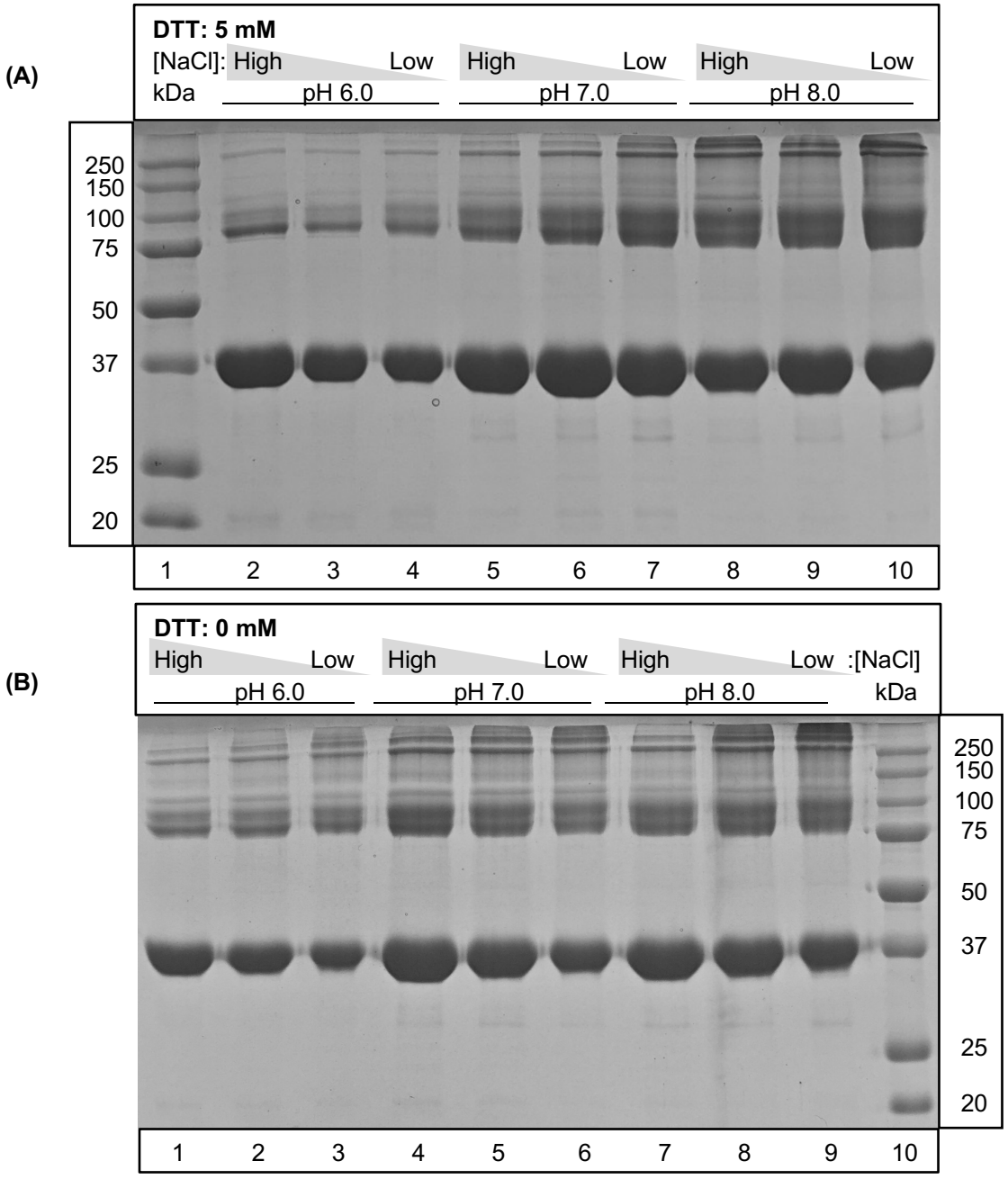


Figure 13. 12% SDS-PAGE gels showing WT *At* ProXp-ala stocks stored at 4°C for 14 days in different pH and salinity buffer conditions and a highly **(A)** or poorly **(B)** reducing environment. All nine conditions are listed in the same order from left to right on each gel.

Impact of At ProXp-ala Truncation on Structure and Oligomerization

Once purification was completed and the concentration of each protein stock determined via Bradford assay, Jun-Kyu Byun performed CD spectroscopy to assess any changes in overall secondary structure that occurred upon truncation. The previously mentioned, computational tools (see **Chapter 2**) predicted that the CTD of *At* ProXp-ala was primarily α-helical. The decrease in molar ellipticity signal intensity at 208 and 220 nm observed by CD supports this prediction (**Fig. 14A**). The general similarity between the two curves suggests that the secondary structures within the N-terminal domain were not severely impacted upon truncation (**Fig. 14A**).

Using purified WT and ΔCTD *At* ProXp-ala, Dr. William Cantara performed duplicate SEC-MALLS experiments to probe the different homo-oligomeric states that each protein formed in solution (**Fig. 14B**). The spectra indicate that WT *At* ProXp-ala adopts several oligomeric states, ranging from a dimer to a hexamer, with a dominant trimer and almost no monomer. In contrast, the spectrum for the ΔCTD indicates no oligomerization activity, as ΔCTD elutes exclusively as a monomer, suggesting that the CTD does indeed possess the same oligomerization capability as the helical region of the *Ct* Pgp3 antigen. However, a single strong trimeric peak was not observed, as would have been expected from clear homo-trimerization activity, so it is currently unclear which of these oligomeric states may be the most relevant *in vivo*. Once more is known about the interacting partners of WT ProXp-ala in vivo, more experiments will be performed to probe the degree to which this oligomerization activity may be modulated by other interactions.

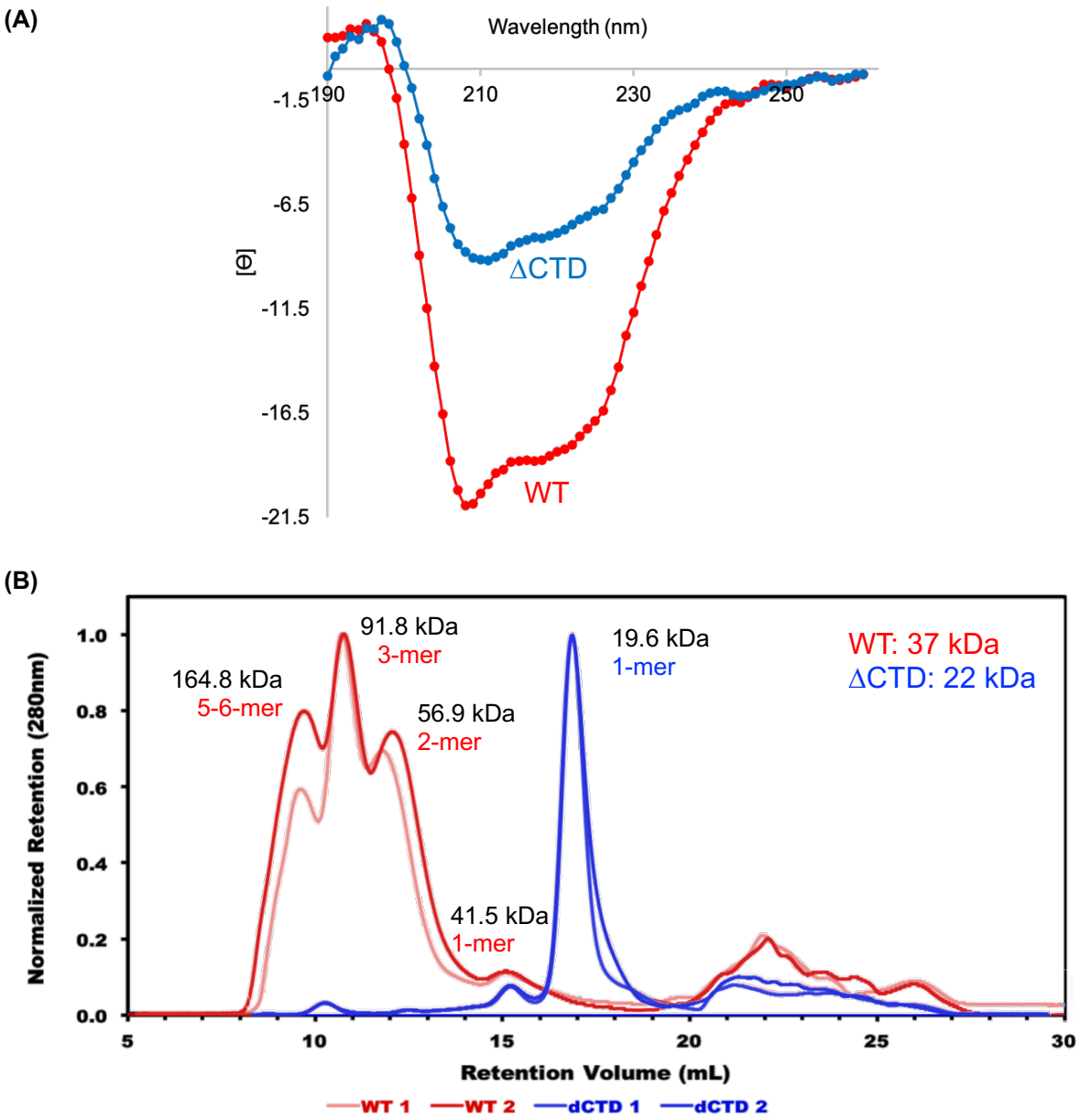


Figure 14. CD and SEC-MALLS spectra comparing WT and ΔCTD *At* ProXp-ala. **(A)** CD spectra of WT (red) and ΔCTD (blue) *At* ProXp-ala. α-helices give high signal intensity at 208 and 220 nm. **(B)** SEC-MALLS spectrum of WT (red) and ΔCTD (blue) *At* ProXp-ala indicating oligomeric states consistent with the molecular weights calculated based on the calibrated retention volume.

Acknowledgements

I would like to thank Dr. Lin Chen for designing the ProXp-ala and tRNA plasmids, Jun-Kyu Byun for generating the ΔCTD mutant protein construct and the CD data, and Dr. Willliam Cantara for his SEC-MALLS data. Thank you to Dr. Marina Bakhtina, Shuohui Liu, and Danni Jin for their help using the FPLC, and to Irina Rouzina for her expertise on protein and tRNA purification. Lastly, I would like to thank Dr. Lexie Kuzmishin Nagy for the extensive technical and theoretical training that she provided me during my first year in lab. I performed most of the aforementioned techniques for the first time under her guidance, and her mentorship set the tone for my career in research and helped me set high standards for myself that have allowed my project to progress.

Chapter 4: Characterization of *Arabidopsis thaliana* Deacylation Kinetics *in vitro*

Introduction

The established *in vitro* function of ProXp-ala in prokaryotes and humans is deacylation of mischarged Ala-tRNA^{Pro}. Briefly, ProRSs present in all domains of life mischarge Ala onto cognate tRNA^{Pro}. In some bacteria, like *Ec*, an editing domain is encoded within the synthetase, but other bacteria, including *Cc*, encode a separate *trans*-editing protein known as ProXp-ala. It is known that *Cc* ProXp-ala functions as a trans-editing deacylase along with *Cc* YbaK to clear Ala-tRNA^{Pro} and Cys-tRNA^{Pro}, respectively[14,23]. However, YbaK sequences are not found in eukaryotes, which raises questions as to how eukaryotic systems manage the challenges of translational fidelity. It has been found that *Hs* ProXp-ala deacylates *Hs* Ala-tRNA^{Pro} much more slowly that *Cc* ProXp-ala deacylates *Cc* tRNA^{Pro}, suggesting that there may have been less selective pressure for robust translational fidelity in humans than in bacteria. Alternatively, other factors present *in vivo* that are missing in our *in vitro* system may facilitate human ProXp-ala deacylation. The difference between prokaryotic and eukaryotic mechanisms has sparked greater interest in studying translational fidelity in other kingdoms like plants, fungi, and protists.

Sequence alignments of ProXp-ala revealed that a CTD exclusive to plants could give some insight into eukaryotic translational fidelity, and comparisons of the deacylation ability of *At* ProXp-ala to that of *Cc* and *Hs* are of particular interest as it is not known to what degree the CTD may participate and influence this process. Since ProXp-ala has never been studied in plants to date, our first functional experiments were to confirm that *At* ProXp-ala exhibits the same canonical function as other ProXp-ala enzymes, then probe how the unique plant-specific CTD impacts that function.

Single-turnover kinetic deacylation assays revealed robust *At* Ala-tRNA^{Pro} deacylation by the WT enzyme and a significant defect in deacylation activity upon truncation of the CTD. Our results are consistent with a binding defect and prompted further study to quantify the difference in affinity via microscale thermophoresis (MST). MST measures the strength of the interaction between two binding partners, one of which is fluorescently labeled, by monitoring the change in fluorescence that occurs when the labeled molecule moves through a temperature gradient induced by an IR laser. Temperature changes and binding interactions change the energetic environment of the fluorophore, resulting in a change in signal. Additionally, the solvation entropy of the bound complex is different from the free protein, resulting in a difference in the rate at which the two species move when trying to reach equilibrium within the applied thermal gradient. All of these effects manifest as an MST signal that can be measured by comparing the fluorescence before and after the application of the thermal gradient. If binding events are observed, plotting the change in fluorescence against substrate concentration on a log scale should generate a sigmoidal curve that can be used to calculate the strength of the interaction via the dissociation constant, K_d.

Another set of deacylation experiments was performed using human tRNA^{Pro} and human ProXp-ala (purified using appropriate sequences in the same vectors and cell lines used for the A*t* constructs) to make preliminary comparisons regarding the degree of functional conservation of ProXp-ala deacylation activity among eukaryotes. These results led to new hypotheses regarding the evolutionary origins of the *At* ProXp-ala CTD.

Materials and Methods

3'-[^{32}P] Radiolabeling of tRNA^{Pro}

A 100 μL reaction containing 1 μM purified *At* tRNA^{Pro} mixed with 2.7 μM purified nucleotidyltransferase (NTase) and 25 μL radioactive α-^{32}P-ATP (Perkin Elmer, ~0.50 mCi, 25 μL) in labeling buffer (50 mM glycine pH 9.0, 10 mM $MgCl_2$, 50 μM $NaPP_i$) was incubated at 37°C. After 5 min, the last two components of the 100 μL reaction were added (10 μM pyrophosphatase (PP_iase) and 1 μM CTP) and the reaction was incubated at 37°C for 2 min. This protocol selectively exchanges the 3' adenosine of *At* tRNA^{Pro} with a radioactively labeled adenosine, generating radiolabeled 3'-[^{32}P]-tRNA^{Pro30}.

To separate the NTase from the tRNA after the reaction , 100 μL acid-phenol:chlorophorm (Ambion) was vortexed with the solution, then spun in a benchtop microcentrifuge (13,300 rcf, 1 min, 4°C). The phenol layer was then removed, and this process was repeated. To remove the excess free α-^{32}P-ATP, Quick Spin Sephadex G-25 Columns were prepared by two rounds of centrifugation (1,900 rcf, 1 minute, 4°C). The top, aqueous layer of the acid-phenol:chloroform extraction was loaded onto the column and the ~75 μL radioactive flow-through was collected.

The tRNA was precipitated via overnight storage at -80°C in 200 μL of cold 100% ethanol and 10 μL of 3 M sodium acetate (pH 5.2). The next day, the sample was centrifuged (13,300 rcf, 30 mins, 4°C) and the ethanol buffer was decanted into a hot sink, leaving behind a white tRNA pellet that was washed in 300 μL of 70% (v/v) ethanol, centrifuged (13,300 rcf, 10 mins, 4°C), decanted again, and finally left to air dry for an hour. The dried pellet was then resuspended in 10 μL of sterilized, filtered, MilliQ water, and stored at -80°C.

At tRNA^{Pro} Mischarging with Alanine using A. thaliana ProRS

Since *At* ProRS naturally mischarges tRNA^{Pro} with Ala when there is no competition with cognate Pro, and it lacks a *cis*- or *trans*-editing domain to deacylate this substrate, *At* ProRS is an optimal way to effectively generate substantial quantities of mischarged *At* Ala-tRNA^{Pro}. To generate Ala-tRNA^{Pro} substrate, 50 μM cold tRNA^{Pro} was mixed with 10 μL of radiolabeled tRNA (concentration of the radiolabeled tRNA < 2 μM, so the radioactivity levels of the sample were considered to represent 50 μM tRNA). The tRNAs were then refolded with a 2 min incubation at 80°C, followed by a 2 min incubation at 60°C, the addition of $MgCl_2$ (10 mM final concentration), and finally a 5 min incubation at 25°C.

The 200 μL mischarging reaction was performed in the following conditions: 10 μM *At* ProRS (purified by Lin Chen), 5 μM 3'-[^{32}P]-radiolabeled *At* tRNA^{Pro} mix, and 1 M Ala, incubated for 10 min at 37°C in mischarging buffer (50 mM HEPES pH 7.5, 20 mM KCl, 20 mM β-ME, 10 mM $MgCl_2$, 4 mM ATP, 100 μM BSA, 6.25% v/v PP_iase). The reaction was quenched with an equal volume acid-phenol:chloroform before it was vortexed and centrifuged (13,300 rcf, 1 min, 4°C) to separate the aqueous and phenol layers. The phenol layer was then discarded, and the extraction was repeated. After the second separation, the mischarged tRNA in the aqueous layer was collected and precipitated in 400 μL of cold 100% ethanol and 20 μL of 3 M sodium acetate (pH 5.2) overnight at -80°C. The next day, the samples were centrifuged (13,300 rcf, 45 mins, 4°C) and the ethanol buffer was removed. The resulting pellet was washed in 400 μL of 70% (v/v) ethanol, centrifuged (13,300 rcf, 20 mins, 4°C), decanted again, and finally left to air dry for an hour. The dried pellet was resuspended in 3 mM sodium acetate (pH 5.2) and stored at -80°C.

To determine the concentration of Ala-tRNA^{Pro} and the degree of aminoacylation, an S1 nuclease (ThermoFisher) digestion was performed to cleave off the 3'-[^{32}P]-AMP of the tRNA. A total of 70 μL of S1 nuclease quenching solution (1600 U S1 nuclease, 1.3 mM S1 nuclease buffer, 130 mM sodium acetate pH 5.2) was prepared using ThermoFisher reagents. 6 μL of S1 quenching solution was combined in duplicate with 2 μL of tRNA^{Pro} before or after the aminoacylation reaction, and the solutions were left to incubate at room temperature for 30 min.

To separate mischarged Ala-[^{32}P]-AMP from uncharged free 3'-[^{32}P]-AMP, the reactions were spotted in triplicate on polyethyleneimine-cellulose thin layer chromatography (TLC) plates (MilliporeSigma). Once the spots dried, the plates were placed in TLC chambers and separated with a mobile phase of ammonium acetate TLC plate running buffer (100 mM ammonium acetate, 5% v/v acetic acid) until the solvent front was 1 cm from the top of the plate. The plate was removed from the chamber and allowed to air dry before being wrapped in clear film and placed in a photosensitive cassette overnight. The next day, the radioactive TLC plate was removed, and the exposed cassette was imaged with a Typhoon phosphorimager. Ala-[^{32}P]-AMP migrates farther up the plate than uncharged free 3'-[^{32}P]-AMP, so ImageQuant densitometry software was used to measure the pixel density of each radioactive spot on the TLC plate image. The degree of charging was determined by taking the pixilation volume of the higher Ala-[^{32}P]-AMP spot over the sum of both the charged and uncharged spots. The concentration was determined by comparison of the sum of both spots to the pixel density of the 5 μM uncharged 3'-[^{32}P]-AMP spots.

Single-Turnover Deacylation Kinetics Assays

Under single-turnover conditions, enzyme concentration is much greater than substrate concentration, so it can be assumed that all of the substrate will be bound to enzyme at the start of the reaction. These conditions were chosen over traditional multiple-turnover experiments, where substrate concentration exceeds enzyme concentration, in part, because of the relatively rapid buffer hydrolysis of Ala-tRNA^{Pro}. The k_{obs} determined by single-turnover kinetics incorporates both the binding and catalysis enzymatic steps and allows direct comparison of WT and ΔCTD ProXp-ala while minimizing the amount of radioactive tRNA necessary for the reaction.

Each assay was performed as a time-course experiment under single-turnover conditions containing 750 nM of either WT or ΔCTD *At* ProXp-ala and 100 nM 3'-[^{32}P]-labeled *At* Ala-tRNA^{Pro} in deacylation assay buffer (50 mM HEPES pH 7.0, 20 mM KCl, 5 mM $MgCl_2$, 2 mM DTT, 0.1 mg/mL BSA). To measure deacylation activity, a 1 μL aliquot was withdrawn from the reaction mixture at the desired timepoints and quenched with the previously described S1 nuclease quenching solution to cleave the radiolabeled 3' adenosine, resulting in a mix of radiolabeled AMP that is covalently bound to Ala or unbound depending on whether it was successfully deacylated. Charged Ala-[^{32}P]-AMP was separated from deacylated [^{32}P]-AMP via spotting on TLC plates as described previously. Photosensitive cassettes and ImageQuant densitometry software were used to measure the pixel density of each radioactive spot on the TLC plate image.

The extent of deacylation was determined by plotting the fraction of Ala-tRNA^{Pro} remaining as a function of time. The k_{obs} for each reaction was determined by fitting the deacylation data to an exponential decay function in the form of $Y= Y_0 + Ae^{-bx}$, where Y = amount of remaining Ala-tRNA^{Pro}, Y_0 = amount of remaining Ala-tRNA^{Pro} after complete deacylation, A = initial amount of Ala-tRNA^{Pro}, b = first order rate constant (k_{obs}), and x = time.

Deacylation assays were performed to confirm activity of the WT and ΔCTD enzymes and identify any potential differences between the two proteins as a result of truncation. Additional deacylation assays under identical conditions, but with a range of enzyme concentrations (0.5-16 μM) were performed by Jun-Kyu Byun (WT) and John Vu (ΔCTD) to probe whether the observed decrease in deacylation activity of the ΔCTD was caused by a catalytic or binding defect. Preliminary assays under identical conditions comparing the deacylation of *Hs* Ala-tRNA^{Pro} by *Hs*, WT *At*, and ΔCTD *At* ProXp-ala were also performed to probe relative levels of activity between eukaryotic ProXp-ala.

Microscale Thermophoresis Binding Assays

For microscale thermophoresis (MST) binding studies, the surface amines of Lys residues on WT and ΔCTD *At* ProXp-ala were fluorescently labeled using the Monolith Protein Labeling Kit RED-NHS 2nd Generation protocol (NanoTemper). The succinimidyl ester fluorophore was reconstituted in 25 μL DMSO to yield a 600 μM dye stock. Each protein was diluted to 10 μM in 90 μL NanoTemper Labeling Buffer NHS. A sample of 7 μL of dye solution was diluted 2X in 7 μL of Labeling Buffer NHS, then 10 μL of this 300 μM dye solution was added to the 90 μL of 10 μM protein for a total reaction volume of 100 μL and a dye to protein reaction stoichiometry of just over 3:1. The reaction was incubated at 25°C in the dark for 30 min, before being loaded onto a NanoTemper Labeling B-Column equilibrated with 10 mL of MST buffer (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% (v/v) Tween-20). MST buffer (550 μL) was added to the column to allow the proteins to enter the resin, then 450 μL more were added to elute the labeled protein, which was collected.

To determine the concentration and degree of labeling of each protein, UV-Vis absorbances were taken at 205, 280, and 650 nm. Since *At* ProXp-ala lacks Trp residues, protein concentration was determined using absorbance of the amide backbone at 205 nm (A_{205}) with correction for the absorbance of the fluorophore at this wavelength. This corrected absorbance was divided by the MW of the protein in Daltons and multiplied by 31 Da^{-1} cm^{-1} according to the NanoTemper labeling protocol. The concentration of label was determined by dividing A_{650} by ε_{650} for the fluorophore (195,000 M^{-1} cm^{-1}). To determine the degree of labeling, the concentration of label was divided by the protein concentration.

For the MST binding assays, a 12-step (WT) or 16 step (ΔCTD) 2X serial dilution of uncharged, unlabeled *At* $tRNA^{Pro}$ was performed ranging from 0-200 μM. Binding equilibrium samples were prepared using 10 μL of the appropriate tRNA dilution combined with 10 μL of 10 nM WT or ΔCTD *At* ProXp-ala for a final protein concentration of 5 nM. The mixtures were allowed to equilibrate at 25°C for approximately 30 min, then they were drawn into Monolith NT.115 Premium Capillaries and measured and analyzed by a Monolith Pico MST instrument.

Results and Discussion

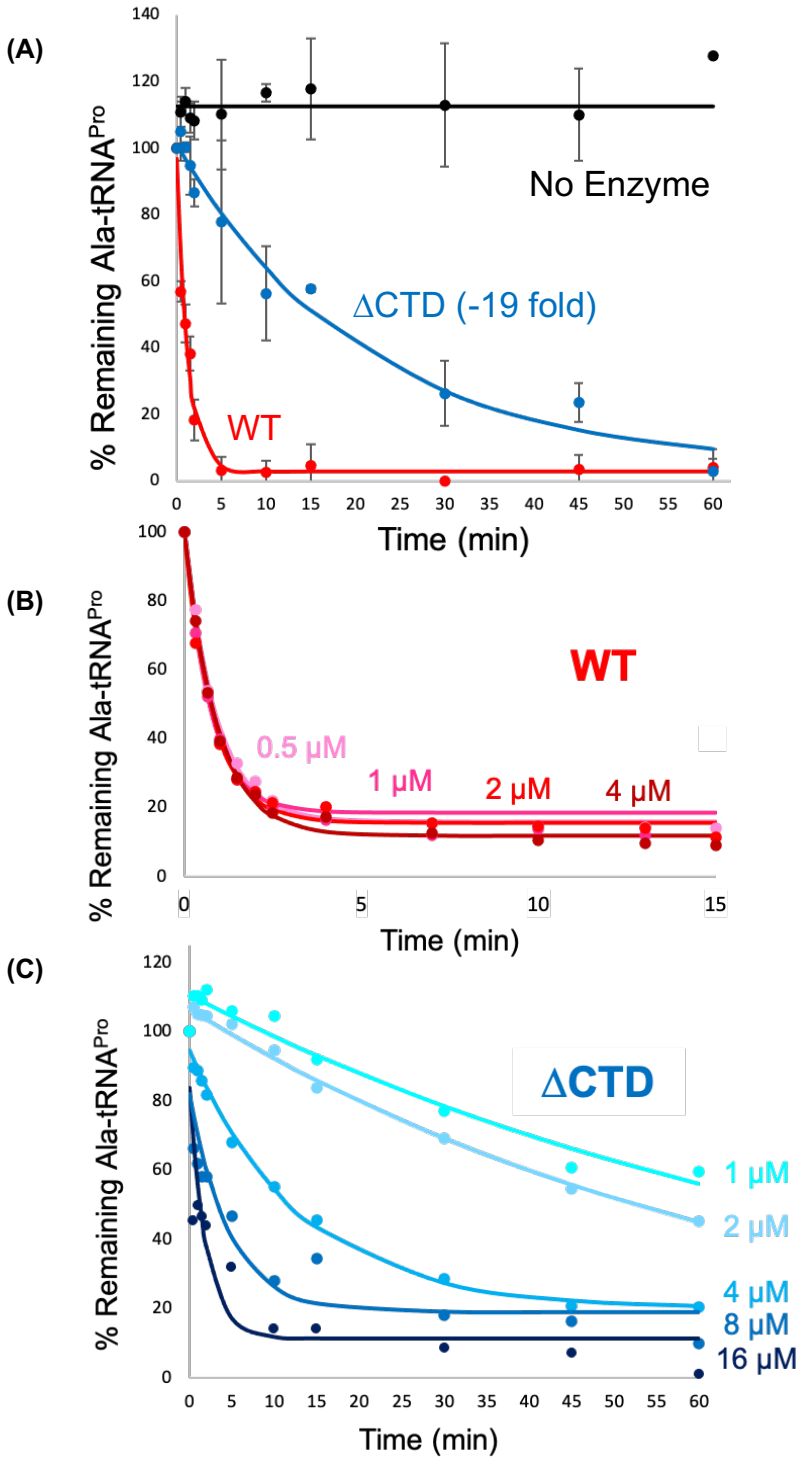
Deacylation Assays

Deacylation assays indicate that CTD truncation results in a 19-fold average decrease in k_{obs}, the apparent first-order rate constant of deacylation activity (**Fig. 15A**). Under single-turnover conditions, the k_{obs} reflects both the binding and catalysis steps. To determine whether the decreased deacylation activity was primarily due to a binding or catalytic defect, deacylation assays were performed with increasing concentrations of ProXp-ala (**Fig. 15B-C**). If the CTD was critical for catalysis, the truncation would limit the enzyme's ability to deacylate Ala-

tRNA^{Pro} independent of how effectively it can bind, which would be a function of enzyme concentration. Therefore, increasing the concentration of a "catalytically defective" ΔCTD ProXp-ala would result in more bound substrate, but the rate of deacylation will be limited by catalysis and the reaction rate would not increase. In contrast, if the enzyme acquired a binding defect from the truncation, providing a higher concentration of the ΔCTD ProXp-ala would increase the amount of bound substrate and would result in an increase in deacylation activity. When the concentration of the ΔCTD enzyme was increased, the deacylation rate increased steadily, suggesting that binding is the rate limiting step for this enzyme (**Fig. 15C**). This observation was not seen in the WT, which exhibited no change in deacylation rate with concentration (**Fig. 15B**). These observations led us to hypothesize that the CTD of plant ProXp-ala greatly improves the binding of the protein to tRNA rather than improving catalysis.

Interspecies tRNA Deacylation

Eukaryotic tRNA^{Pro} sequences are very similar to one another (**Fig. 16A**), so we hypothesized that *At* ProXp-ala would be able to successfully deacylate mischarged *Hs* Ala-tRNA^{Pro}. Deacylation assays of WT *At*, ΔCTD *At*, and *Hs* ProXp-ala using mischarged *Hs* Ala-tRNA^{Pro} were performed under single turnover conditions and their rates of deacylation were compared (**Fig. 16B**). As hypothesized, both plant enzymes were able to successfully deacylate *Hs* Ala-tRNA^{Pro}, and the difference between their k_{obs} was about 22-fold, very similar to the 19-fold difference seen with *At* Ala-tRNA^{Pro} (**Fig. 16B**). The human enzyme, which lacks a CTD, exhibited the poorest deacylation activity of the three enzymes tested, deacylating its human substrate about 2-fold slower than ΔCTD *At* ProXp-Ala (**Fig. 16B**).



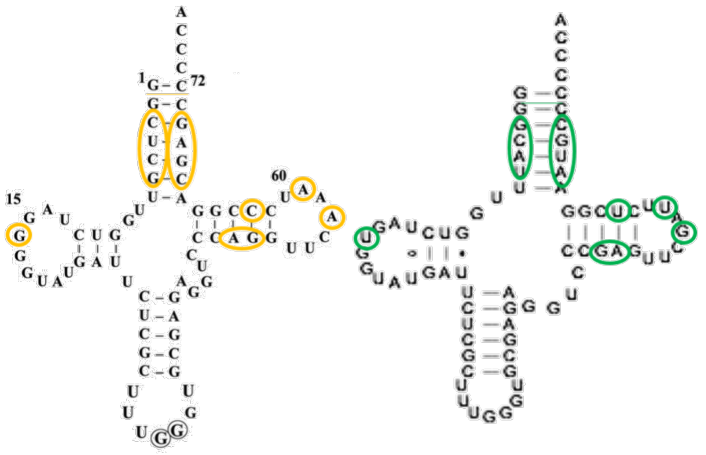
Enzyme	k_{obs} (min^{-1})
WT	0.805 ± 0.066
ΔCTD	0.043 ± 0.014

[WT]	k_{obs} (min^{-1})
0.5 μM	1.13
1 μM	1.36
2 μM	1.23
4 μM	1.09

[ΔCTD]	k_{obs} (min^{-1})
1 μM	0.012
2 μM	0.014
4 μM	0.077
8 μM	0.215
16 μM	0.506

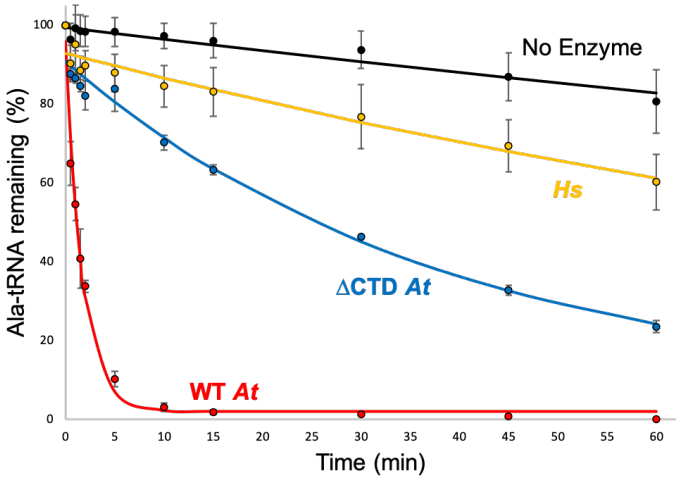
Figure 15. Single-turnover kinetic assays of **(A)** 750 nM WT and ΔCTD *At* ProXp-ala deacylation of *At* Ala-tRNA^{Pro} performed in duplicate with standard deviations indicated. Single trials of deacylation assays at varying enzyme concentrations for **(B)** WT (performed by Jun-Kyu Byun) and **(C)** ΔCTD *At* ProXp-ala. Observed rate constants, k_{obs}, were determined by fitting the kinetic data to a first-order exponential decay function. Curves represent fits of the data to this equation.

(A)



***H. sapiens* tRNA^{Pro}** **_A. thaliana_ tRNA^{Pro}**

(B)



(C)

ProXp-ala Variant	K_{obs} (min^{-1})
WT *H. sapiens*	0.015 ± 0.012
ΔCTD *A. thaliana*	0.025 ± 0.003
WT *A. thaliana*	0.591 ± 0.083

Figure 16. **(A)** Sequence and secondary structure of *Hs* and *At* tRNA$^{Pro(UGG)}$. Positions that are not conserved between the two species are circled. **(B)** Single-turnover kinetic assays of WT and ΔCTD *At* ProXp-ala and *Hs* ProXp-ala deacylation of *Hs* Ala-tRNA^{Pro} performed in duplicate with the standard deviation indicated. **(C)** k_{obs} values and first-order exponential decay curves for each dataset were determined as described in **Fig 15**.

ΔCTD *At* ProXp-ala was designed to mimic the size, structure, and function of shorter eukaryotic ProXp-ala, such as the protein found in humans. The similarity in their deacylation activity suggests that the ΔCTD variant may to some degree model the ancestral plant ProXp-ala from around the time that plants diverged from other eukaryotes before the evolution of a CTD. It is also worth noting that these findings indicate that, rather than the CTD performing a compensatory function to recover similar levels of deacylation as other eukaryotes, its rapid deacylation activity has actually improved the activity of plant ProXp-ala beyond the capacity of the catalytic N-terminal core alone and the potentially more ancestral version seen in humans.

Comparative Binding Affinities

Preliminary MST experiments in MST buffer (50 mM Tris-HCl pH 7.8, 150 mM NaCl, 10 mM $MgCl_2$, 0.05% v/v Tween-20) were performed with 5 nM of the WT and ΔCTD proteins. The percent labeling with the Lys-specific fluorophore used for MST was inconsistent between the two proteins and will require further optimization (WT: 160% and ΔCTD: 7%). Binding checks of both proteins confirmed that a sufficient signal-to-noise ratio was present when 25 μM of tRNA was used. Duplicate trials wherein the tRNA concentration ranged from 0 to 200 μM indicated that the WT protein appeared to exhibit binding behavior, but that saturating concentrations of $tRNA^{Pro}$ were never achieved, even at 200 μM tRNA; thus, a K_d could not be determined for the WT protein (**Fig. 17A**). In contrast, effective saturation of the ΔCTD construct was achieved under the same conditions (but with four more steps in the serial dilution) as for the WT protein, and a binding curve was produced from quadruplicated trials, resulting in a calculated K_d of approximately 840 nM **(Fig. 17B).**

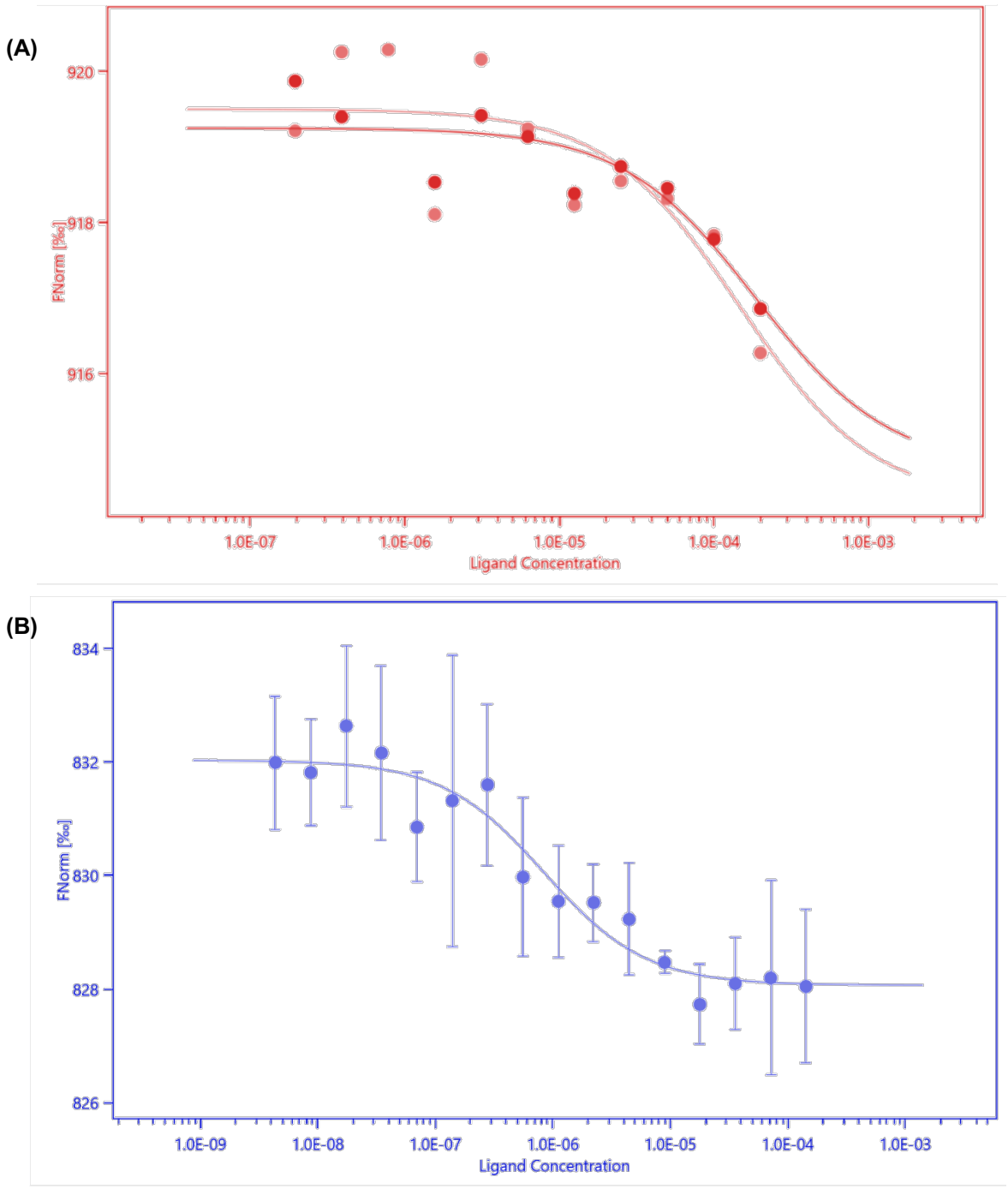


Figure 17. MST binding curves for two separate trials of WT (top) and four combined trials of ΔCTD (bottom). Normalized fluorescence (‰ relative to initial fluorescence) is plotted against molar *At* $tRNA^{Pro}$ concentration on a log scale. WT trials have not reached saturating conditions and cannot be used to determine a K_d. The preliminary K_d of ΔCTD is 840 ± 290 nM.

The MST results suggest that the K_d of WT *At* ProXp-ala exceeds that of ΔCTD ProXp-ala, a result that is inconsistent with our previously proposed model of the CTD facilitating binding to tRNA. While our results are preliminary and require optimization before conclusions can be drawn with greater confidence, a possible explanation for this result is that a longer time to equilibration may be necessary for the WT protein. Incubation for the first trials was only 30 minutes, and longer incubation times may be necessary to achieve true equilibrium binding measurements[31]. Alternatively, the labeling of the fluorophore may be disrupting the function of the Lys-rich CTD, preventing any crucial protein-protein and protein-tRNA interactions. To assess the potential effect that labeling could be having on protein function, a control deacylation assay will need to be performed to ensure that fluorescently labeled WT and ΔCTD *At* ProXp-ala show similar changes in deacylation activity upon labeling. If the WT exhibits a more significant drop in deacylation activity after surface Lys labeling, or fails to exhibit the expected oligomerization seen in unlabeled WT, then alternative labeling methods must be considered to be able to continue performing MST.

Acknowledgements

Thank you to Dr. Lexie Kuzmishin Nagy for her radioactivity training and for teaching me the rapid pipetting techniques necessary for deacylation assays. Special thanks to Dr. Miljan Simonović, Anu Puppala, and Kaitlyn Kiernen Powers at the University of Illinois-Chicago for their mentorship and introduction to microscale thermophoresis (MST) and for welcoming me so generously into their research space during the summer of 2019. Thank you also to Dr. Dmitri Kudryashov for allowing me to use his lab's Monolith instrument to continue the MST experiments at Ohio State.

Conclusions and Future Directions

With this work, we have confirmed the expected canonical function of *At* ProXp-ala as an efficient Ala-tRNA^{Pro} deacylase and provided insights into the structure and function of this protein's enigmatic C-terminal domain. Kinetic deacylation assays demonstrate that truncation of the *At* ProXp-ala CTD results in a 19-fold decrease in deacylation rate. The rate can be improved by increasing the concentration of enzyme in the reaction, a result that is consistent with a model in which the CTD improves the binding affinity of ProXp-ala to Ala-tRNA^{Pro}. Pilot electrophoretic mobility shift assays performed by Jun-Kyu Byun point to a binding defect as well, though attempts using microscale thermophoresis as an alternative technique to quantify the difference in the strength of binding between the *At* ProXp-ala variants and *At* tRNA^{Pro} are not yet conclusive. More experiments and optimization are underway to provide conclusive results regarding this parameter, though it is clear that the CTD has some critical function that manifests as a change in deacylation rate.

SEC-MALLS measurements performed by Dr. William Cantara from the Musier-Forsyth lab are consistent with the WT protein forming a higher order dimer or trimer, whereas the ΔCTD variant is exclusively monomeric. Additionally, SEC-MALLS measurements performed by Dr. Marina Bakhtina with the WT protein and *At* tRNA^{Pro} revealed that WT ProXp-ala forms a complex with tRNA exclusively as a dimer (data not shown). These data are consistent with bioinformatics analyses (Chapter 2) and suggest that oligomerization may be a pre-requisite for effective tRNA binding. Further studies investigating the binding stoichiometry of the ΔCTD to tRNA are planned, as well as experiments to probe these interactions via native mass spectrometry.

Given the structural and functional similarities between human ProXp-ala and ΔCTD *At* ProXp-ala, it is very likely that the CTD evolved uniquely in plants from a common eukaryotic ProXp-ala ancestor similar to the human ProXp-ala. Once the functions of the CTD are better characterized, we will design a chimeric +CTD *Hs* ProXp-ala protein construct in which the CTD of *At* ProXp-ala is fused to the C-terminus of *Hs* ProXp-ala. Because of the similarity in the structure and function of *Hs* and ΔCTD *At* ProXp-ala, and the evidence that has been found for the CTD performing a facilitatory binding function, we hypothesize that this chimeric fusion would exhibit similarly improved (~20 fold) deacylation activity compared to the WT *Hs* enzyme without a CTD. Improved deacylation by the chimera will suggest an ancient evolutionary mechanism that placed selective pressure on plants to improve the rate of ProXp-ala deacylation activity that was not present in human evolution. This deficiency in ProXp-ala deacylation within humans, if confirmed *in vivo*, could therefore serve as a potential limitation of human translational fidelity machinery, and it could inform potential mechanisms of disease, such as neurodegeneration which has been shown to be a product of failed aa-tRNA editing[4].

The tendency for WT At ProXp-ala to interact with itself during storage has also informed future plans that involve looking deeper at the types of protein-protein interactions that may be facilitated by the CTD *in vivo*. In fact, while it has been well established that humans possess a multi-synthetase complex (MSC), very recently, the first plant MSC was identified, and it was found that *At* ProXp-ala participates in this complex[32]. This finding is in contrast to *Hs* ProXp-ala which has not been reported to participate in the human MSC. Therefore, we hypothesize that the CTD may play a significant role in the ability of *At* ProXp-ala to stably associate with the plant MSC. Thus, comparing the *in vivo* binding partners of WT and ΔCTD *At* ProXp-ala is a future direction of this project.

This work has begun to shed a light on the structure, function, and evolution of the enigmatic C-terminal extension of plant ProXp-ala. Our current results are consistent with a model in which the ubiquitously conserved N-terminal catalytic core of the protein retains a level of Ala-tRNA^{Pro} deacylase activity similar to other eukaryotic systems, while a highly helical CTD attached to the N-terminal core by a random linker domain appears to facilitate a greater rate of *in vitro* deacylation, potentially by enhancing binding to the Ala-tRNA^{Pro} substrate or facilitating protein oligomerization that improves deacylation.

Our findings have expanded our understanding of eukaryotic translational fidelity to the plant kingdom and, in doing so, have revealed alternative mechanisms of translational fidelity between eukaryotic systems. These findings have potential industrial implications for pesticides by exploiting differences in the efficiency of translational fidelity machinery between plants, parasitic organisms, and bacteria. If plants are found to have enhanced translational fidelity relative to other organisms *in vivo*, then plants may be tolerant to higher levels of proteinogenic and nonproteinogenic amino acids that would be toxic to other species. The results of this study may also inform future medical applications by identifying a potential mechanism for overcoming poor translational fidelity in humans, which has been implicated in neurodegeneration and cardioproteinopathy in animal models[4,33].

References

1 Yadavalli, S. S. and Ibba, M. (2012) Quality control in aminoacyl-tRNA synthesis its role in translational fidelity. Adv Protein Chem Struct Biol 86, 1-43, doi:10.1016/b978-0-12-386497-0.00001-3

2 Arnez, J.G. and Moras, D. (1997) Structural and functional considerations of the aminoacylation reaction. Trends Biochem. Sci. 22 (6) 211–216

3 Nangle, L. A., Motta, C. M. and Schimmel, P. (2006) Global effects of mistranslation from an editing defect in mammalian cells. Chem Biol 13, 1091-1100, doi:10.1016/j.chembiol.2006.08.011

4 Lee, J. W., Beebe K., Nangle, L. A., Jang, J., Longo-Guess, C. M., Cook, S. A., Davisson, M. T., Sundberg, J. P., Schimmel, P., and Ackerman, S. L. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature. 2006 Sep 7;443(7107):50-5. doi: 10.1038/nature05096. Epub 2006 Aug 13. PMID: 16906134.

5 Ibba, M. and Soll, D. (2000) Aminoacyl-tRNA synthesis. Annu Rev Biochem 69, 617-650, doi:10.1146/annurev.biochem.69.1.617

6 Mascarenhas, A. P., An, S., Rosen, A. E., Martinis, S. A., and Musier-Forsyth, K. Fidelity (2009) Mechanisms of the Aminoacyl-tRNA Synthetases. Protein Engineering: Nucleic Acids and Molecular Biology 22 155-203, doi:10.1007/978-3-540-70941-1_6

7 Ibba, M. and Söll, D. (1999) Quality control mechanisms during translation, Science 286 (5446) 1893–1897.

8 Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res 26, 5017-5035

9 Pauling, L. (1957) The Probability of Errors in the Process of Synthesis of Protein Molecules, Vol, Birkhauser, Basel

10 Ling, J., Reynolds, N., and Ibba, M. (2009) Aminoacyl-tRNA synthesis and translational quality control. Annu Rev Microbiol 63, 61-78, doi:10.1146/annurev.micro.091208.073210

11 Eldred, E. W. and Schimmel, P. R. (1972) Rapid deacylation by isoleucyl transfer ribonucleic acid synthetase of isoleucine-specific transfer ribonucleic acid aminoacylated with valine. J Biol Chem 247, 2961-2964

12 Baldwin, A. N. and Berg, P. (1966) Transfer ribonucleic acid-induced hydrolysis of valyladenylate bound to isoleucyl ribonucleic acid synthetase. J Biol Chem 241, 839-845

13 Fukunaga, R., and Yokoyama, S. (2005) Structural basis for substrate recognition by the editing domain of isoleucyl-tRNA synthetase, J. Mol. Biol. 359 (4) 901–912

14 Kuzmishin Nagy, A. B., Bakhtina, M., and Musier-Forsyth, K. (2020) *trans*-Editing by aminoacyl-tRNA synthetase-like editing domains. L. Ribas de Pouplana, L. Kaguni (Eds.), The Enzymes: Biology of Aminoacyl-tRNA Synthetases, Academic Press, Orlando, FL, 1-47

15 Beuning, P. J. and Musier-Forsyth, K. (2000) Hydrolytic editing by a class II aminoacyl-tRNA synthetase. Proc Natl Acad Sci U S A 97, 8916-8920, doi:10.1073/pnas.97.16.8916

16 Ahel, I., Stathopoulos, C., Ambrogelly, A., Sauerwald, A., Toogood, H., Hartsch, T., Söll, D. (2002) Cysteine activation is an inherent in vitro property of prolyl-tRNA synthetases. J Biol Chem 277, 34743-34748, doi:10.1074/jbc.M206928200

17 Splan, K. E., Ignatov, M. E., and Musier-Forsyth, K. (2008) Transfer RNA modulates the editing mechanism used by class II prolyl-tRNA synthetase. J Biol Chem 283, 7128-7134, doi:10.1074/jbc.M709902200

18 Kumar, S., Das, M., Hadad, C. M., and Musier-Forsyth, K. (2012) Substrate specificity of bacterial prolyl-tRNA synthetase editing domain is controlled by a tunable hydrophobic pocket. J Biol Chem 287, 3175-3184, doi:10.1074/jbc.M111.313619

19 Chen, L., Tanimoto, A., So, B. R., Bakhtina, M., Magliery, T. J., Wysocki, V. H., and Musier-Forsyth, K. (2019) Stoichiometry of triple-sieve tRNA editing complex ensures fidelity of aminoacyl-tRNA formation. Nucleic acids research, 47(2), 929–940. https://doi.org/10.1093/nar/gky1153

20 An, S., and Musier-Forsyth, K. (2004) Trans-editing of Cys-tRNAPro by *Haemophilus influenzae* YbaK protein. J Biol Chem 279, 42359-42362, doi:10.1074/jbc.C400304200

21 Ahel, I., Korencic, D., Ibba, M., and Söll, D. (2003) Trans-editing of mischarged tRNAs. Proc Natl Acad Sci U S A 100, 15422-15427, doi:10.1073/pnas.2136934100

22 Crepin, T., Yaremchuk, A., Tukalo, M., and Cusack, S. (2006) Structures of two bacterial prolyl-tRNA synthetases with and without a cis-editing domain. Structure 14, 1511-1525, doi:10.1016/j.str.2006.08.007

23 Vargas-Rodriguez, O. and Musier-Forsyth, K. (2013) Exclusive use of trans-editing domains prevents proline mistranslation. J Biol Chem 288, 14391-14399, doi:10.1074/jbc.M113.467795

24 Guo, M., Yang, X. L., and Schimmel, P. (2010) New functions of aminoacyl-tRNA synthetases beyond translation. Nat Rev Mol Cell Biol. 2010 Sep;11(9):668-74. doi: 10.1038/nrm2956. Epub 2010 Aug 11. PMID: 20700144; PMCID: PMC3042954

25 Kim, M. H., and Kim, S. (2020) Chapter Six - Structures and functions of multi-tRNA synthetase complexes, L. Ribas de Pouplana, L. S. Kaguni (Eds). The Enzymes, Academic Press, 48, 149-173. https://doi.org/10.1016/bs.enz.2020.06.008

26 Mirabello, C. and Pollastri, G. (2013) Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility", Bioinformatics, 29(16):2056-2058, doi: 10.1093/bioinformatics/btt344

27 Chiu, J., March, P. E., Lee, R., and Tillett, D. (2004) Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. Nucleic Acids Res 32, e174, doi:10.1093/nar/gnh172

28 Milligan, J. F. and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. Methods Enzymol.180:51-62. doi: 10.1016/0076-6879(89)80091-6. PMID: 2482430.

29 Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, 72 (1–2) 248-254, https://doi.org/10.1016/0003-2697(76)90527-3

30 Ledoux, S. and Uhlenbeck, O. C. (2008) 3'-[32P]-labeling tRNA with nucleotidyltransferase for assaying aminoacylation and peptide bond formation. Methods 44, 74-80, doi:10.1016/j.ymeth.2007.08.001

31 Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P., and Herschlag, D. (2020) How to measure and evaluate binding affinities. Elife. Aug 6;9:e57264. doi: 10.7554/eLife.57264. PMID: 32758356; PMCID: PMC7452723

32 McWhite, C.D., Papoulas, O., Drew, K., Cox, R. M., June, V., Dong, O. X., Kwon, T., Wan, C., Salmi, M. L., Roux, S. J., Browning, K. S., Chen, Z. J., Ronald, P.C., and

Marcotte, E. M. (2020) A pan-plant protein complex map reveals deep conservation and novel assemblies. Cell. Apr 16;181(2):460-474.e14. doi: 10.1016/j.cell.2020.02.049

33 Liu Y, Satz JS, Vo MN, Nangle LA, Schimmel P, Ackerman SL. Deficiencies in tRNA synthetase editing activity cause cardioproteinopathy. Proc Natl Acad Sci U S A. 2014 Dec 9;111(49):17570-5. doi: 10.1073/pnas.1420196111. Epub 2014 Nov 24. PMID: 25422440; PMCID: PMC4267364